OM of: US-09-303-518D-651 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 7:01 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-651
Query length: 4350
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304	+	1536.50	2047.09	4.6e-107	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564	+	1522.00	2028.28	5.5e-106	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745	+	1056.50	1399.46	6.1e-71	13				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592	+	1038.00	1374.41	1.5e-69	13				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705	+	709.50	945.68	4.9e-45	32				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136	+	385.50	498.91	1.6e-20	74				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842	+	372.50	473.40	2.0e-19	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505	+	282.00	351.28	1.3e-12	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846	+	282.00	351.28	1.3e-12	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293	+	277.00	344.52	3.1e-12	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723	+	271.00	336.41	8.8e-12	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725	+	265.00	328.30	2.5e-11	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824	+	264.50	330.00	2.5e-11	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830	+	264.00	329.37	2.7e-11	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828	+	264.00	329.32	2.7e-11	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825	+	263.50	328.70	3.0e-11	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826	+	261.50	328.38	3.9e-11	97				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827	+	261.00	327.77	4.2e-11	96				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120	+	260.00	317.07	6.9e-11	23				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724	+	259.50	321.28	6.4e-11	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732	+	258.00	318.89	8.4e-11	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609	+	256.50	319.10	1.0e-10	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292	+	256.50	316.37	1.1e-10	16				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01847	+	254.50	317.69	1.4e-10	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01833	+	252.00	315.22	2.0e-10	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01832	+	252.00	315.15	2.1e-10	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837	+	251.00	313.17	2.5e-10	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836	+	251.00	313.11	2.5e-10	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731	+	251.00	310.84	2.7e-10	13				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728	+	251.00	309.79	2.8e-10	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506	+	251.00	309.79	2.8e-10	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848	+	251.00	309.79	2.8e-10	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291	+	251.00	308.96	2.9e-10	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294	+	250.50	309.11	3.1e-10	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG30355	+	242.50	297.41	1.3e-09	16				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143	+	241.50	293.29	1.6e-09	20				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017	+	241.50	282.48	2.4e-09	57				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01842	+	240.50	299.75	1.5e-09	99				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841	+	240.00	299.01	1.6e-09	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840	+	240.00	298.94	1.6e-09	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98256	+	236.00	290.66	3.6e-09					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01843	+	232.50	288.99	6.0e-09					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	+	232.00	283.16	7.8e-09					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834	+	231.00	285.86	8.1e-09					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW04505	+	230.00	278.24	1.2e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69133	+	230.00	278.24	1.2e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69137	+	230.00	278.24	1.2e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69134	+	230.00	277.95	1.2e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01835	+	229.50	283.89	1.0e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR26503	+	229.00	285.16	1.1e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR25578	+	227.00	282.33	1.5e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG24938	+	225.00	277.86	2.3e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:AAR14320	+	224.00	278.41	2.5e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01849	+	223.00	275.69	3.2e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01845	+	222.00	272.96	3.9e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01844	+	222.00	272.89	3.9e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	+	220.50	272.51	4.8e-08					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75097	+	215.50	256.17	1.6e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB46351	+	215.00	253.98	1.8e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37403	+	212.50	242.43	3.7e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365	+	211.50	259.53	2.4e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB66878	+	211.00	249.79	3.5e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE00701	+	209.00	249.17	4.6e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69135	+	208.50	248.85	5.0e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01839	+	207.50	255.58	4.5e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01838	+	207.50	255.51	4.5e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB60656	+	205.50	244.62	8.4e-07					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75096	+	204.50	243.47	1.0e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52855	+	203.50	239.07	1.3e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB62108	+	202.50	247.21	1.1e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69806	+	202.50	242.16	1.3e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB18250	+	202.50	241.88	1.4e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB23860	+	201.00	237.01	1.9e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR99393	+	200.50	236.60	2.1e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB97070	+	200.00	237.34	2.2e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15945	+	200.00	235.79	2.3e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW98828	+	199.00	232.38	2.9e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW71556	+	199.00	232.38	2.9e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR05041	+	195.50	225.20	5.8e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61077	+	194.00	234.94	5.1e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW65088	+	193.50	231.14	6.2e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52738	+	193.50	230.67	6.3e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW42634	+	192.00	226.93	8.6e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95550	+	190.00	232.92	9.1e-06					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75098	+	189.00	220.00	1.6e-05					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU22961	+	188.50	235.57	1.0e-05					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61539	+	188.50	228.64	1.3e-05					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR04232	+	188.50	226.93	1.3e-05					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198	+	188.00	219.46	1.8e-05					

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304

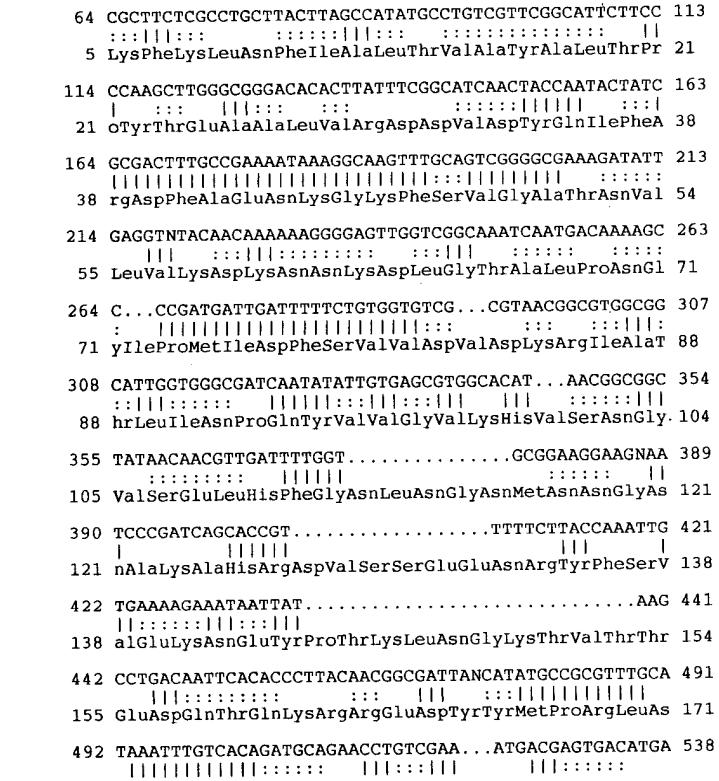
seq_documentation_block:
ID AAR07304 standard; protein; 1541 AA.
XX
AC AAR07304;
XX
DT 31-JAN-1991 (first entry)
XX
DE IgA1 protease.
XX
KW IgA1; vaccine; meningitis; gonnorhoea; allergies.
XX
OS Haemophilus influenzae.
XX
PN WO9011367-A.
XX
PD 04-OCT-1990.
XX
PF 16-MAR-1990; 90WO-DK00073.
XX
PR 17-MAR-1989; 89DK-0001308.
XX

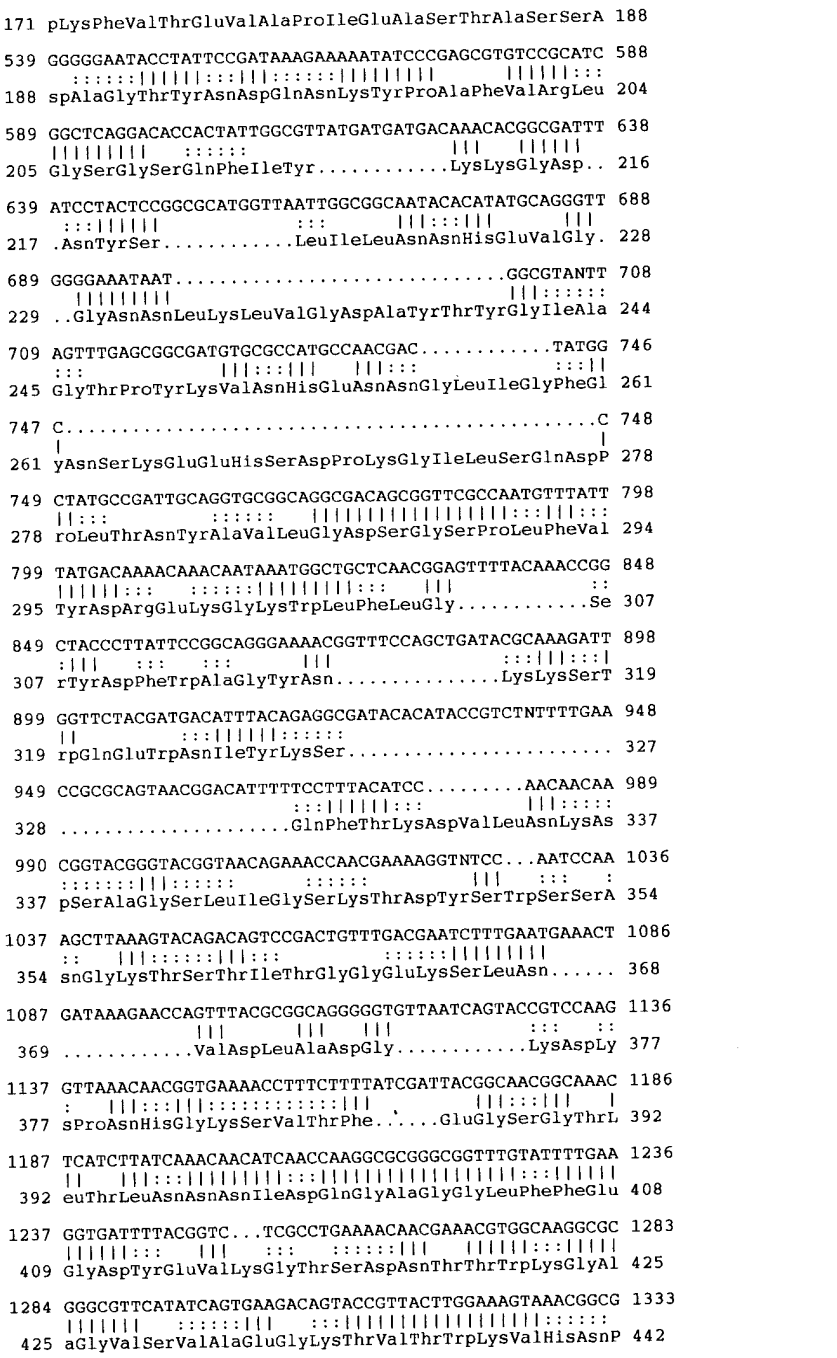
PA (KILI/) KILIAN M.
XX
PI Kilian M, Poulsen K;
XX
DR WPI; 1990-320267/42.
DR N-PSDB; AAQ06164.
XX
PT Immunoglobulin A1 protease prodn. - by cloning from
PT microorganisms for immunisation against immunoglobulin A1
PT protease producing bacteria
XX
PS Disclosure; fig 3; 44pp; English.
XX
CC This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,
CC gonorrhoea or allergic diseases. It specifically cleaves the heavy
CC chain of human IgA1 in the hinge region.
XX
SQ Sequence 1541 AA;

alignment_scores:
Quality: 1536.50 Length: 1722
Ratio: 1.677 Gaps: 56
Percent Similarity: 53.194 Percent Identity: 26.887

alignment_block:
US-09-303-518D-651 x AAR07304 ..

Align seg 1/1 to: AAR07304 from: 1 to: 1541





```
1334 TGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAA 1383
442 roGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 458

1384 GCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCAT 1433
459 GlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyThrValIl 475

1434 TTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAA 1483
475 eLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSerV 491

1484 TCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1533
491 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAspLysGln 507

1534 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTT 1583
508 ValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLeuAspLe 524

1584 AAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGG 1633
524 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspAspGlyA 541

1634 CGATGATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATTACA 1683
541 laArgLeuValAsnHisAsnMetThrAsnAlaSerAsnIleThrIleThr 557

1684 GGGAATGAAAGTATTACACAACCGAGT..... 1710
558 GlyGluSerLeuIleThrAspProAsnThrIleThrProTyrAsnIleAs 574

1710 ..... 1710
574 pAlaProAspGluAspAsnProTyrAlaPheArgArgIleLysAspGlyG 591

1710 ..... 1710
591 lyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyrTyrAlaLeuArgLys 607

1710 ..... 1710
608 GlyAlaSerThrArgSerGluLeuProLysAsnSerGlyGluSerAsnGl 624

1711 .....GGTAAGAAT.....A 1720
624 uAsnTrpLeuTyrMetGlyLysThrSerAspGluAlaLysArgAsnValM 641

1721 TCAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGC 1770
641 etAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly 657

1771 GAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCC 1820
658 GluGluGlu...GlyLysAsnAsnGlyAsnLeuAsnValThrPheLysGl 673

1821 CGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACG 1870
673 yLysSerGluGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 690

1871 GCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA 1920
690 lyAspLeuThrValGluLysGlyThrLeuPheLeuSerGlyArgProThr 706

1921 CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT.. 1968
707 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspProHi 723

1969 ....ATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGCA 2014
723 sPheAlaGluAsnAsnGluValValValGluAspAspTrpIleAsnArgA 740

2015 CGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCC 2064
740 snPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeuTyrSer 756

2065 ...CGCAATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGC 2111
757 GlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAsnLysAl 773

2112 CCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACAC 2161
773 aGlnValHisIleGlyTyr......LysThrGlyAspThrValCysValA 788

2162 GTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGAC 2211
788 rgSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 803

2212 GATAAAGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAG 2261
804 AspLysAlaLeuAsnSerPheAsnProThr..... 813

2262 NCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANG 2311
813 ..... 813

2312 GCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCC 2361
814 ..AsnLeu..... 815

2362 ACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAA 2411
816 ......ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheValLeuGl 830

2412 TCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTA 2461
830 yLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyAsnSerGlnValA 847

2462 ATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAAC 2511
847 rgLeuThrGluAsn..... 851

2512 GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGC 2561
851 ..... 851

2562 CGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCA 2611
851 ..... 851

2612 GCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTG 2661
852 .....SerHisTrpHisLeu 856

2662 CCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTAC 2711
857 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleHi 873

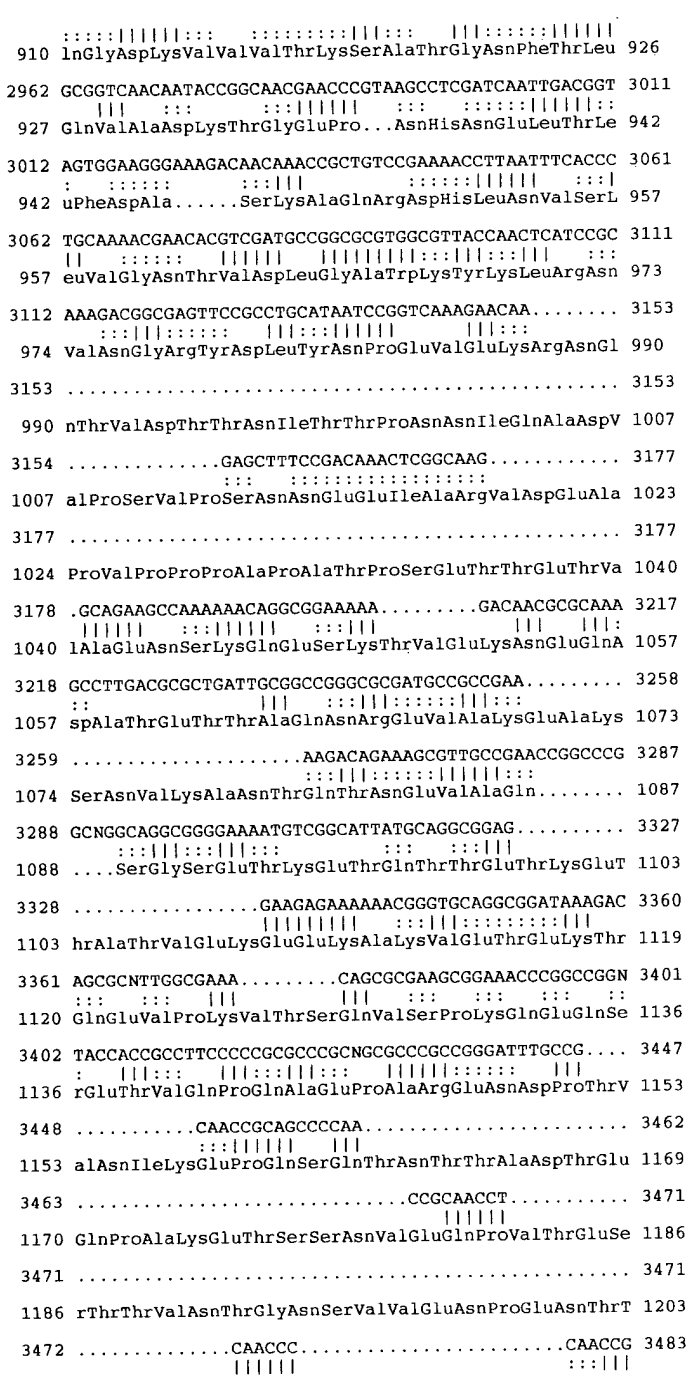
2712 ACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNG 2761
873 sLeuAsnSerAlaAspAsnSerAsn..... 881

2762 TGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACA 2811
882 .....AsnValThr 884

2812 CCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAA 2861
885 .....LysTyrAsnThrLeuThrValAsn...Se 893

2862 ATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACC 2911
893 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 910

2912 GAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTG 2961
```


910 lnGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 926
2962 GCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGT 3011
927 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLe 942
3012 AGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCC 3061
942 uPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnValSerL 957
3062 TGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGC 3111
957 euValGlyAsnThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsn 973
3112 AAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA..... 3153
974 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGl 990
3153 3153
990 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1007
3154 GAGCTTTCCGACAAACTCGGCAAG..... 3177
1007 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1023
3177 3177
1024 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1040
3178 .GCAGAAGCCAAAAAACAGGCGGAAAAA.....GACAACGCGCAAA 3217
1040 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1057
3218 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAA..... 3258
1057 spAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073
3259 AAGACAGAAAGCGTTGCCGAACCGGCCCG 3287
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087
3288 GCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAG..... 3327
1088 SerGlySerGluThrLysGluThrGlnThrThrGluThrLysGluT 1103
3328 GAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3360
1103 hrAlaThrValGluLysGluGluLysAlaLysValGluThrGluLysThr 1119
3361 AGCGCNTTGGCGAAA.....CAGCGCGAAGCGGAAACCCGGCCGGN 3401
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluGlnSe 1136
3402 TACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCG.... 3447
1136 rGluThrValGlnProGlnAlaGluProAlaArgGluAsnAspProThrV 1153
3448 CAACCGCAGCCCCAA..... 3462
1153 alAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1169
3463 CCGCAACCT..... 3471
1170 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluSe 1186
3471 3471
1186 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1203
3472 CAACCC.....CAACCG 3483


1203 hrProAlaThrThrGlnProThrValAsnSerGluSerSerAsnLysPro 1219
3484 CAGCGC..... 3489
1220 LysAsnArgHisArgArgSerValArgSerValProHisAsnValGluPr 1236
3490 GACCTGA 3496
1236 oAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCysAspLeuT 1253
3497 TNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC..... 3537
1253 hrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAlaGln 1269
3538 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGT 3581
1270 PheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1286
3582 G...TTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGN 3628
1286 uGluMetAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrSerM 1303
3629 ACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACC 3678
1303 etAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSer 1319
3679 GACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGT 3725
1320 ThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLe 1336
3726 CGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCA 3775
1336 uGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysAlaT 1353
3776 TCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGC 3825
1353 hrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyrAla 1369
3826 ATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGG 3875
1370 AspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSe 1386
3876 CANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATT 3925
1386 rLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnP 1403
3926 ACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAA 3975
1403 heGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThr 1419
3976 CCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTA 4025
1420 ProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLe 1436
4026 CGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGNG 4075
1436 uAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaPheA 1453
4076 CGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATNTCCATC 4125
1453 laGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheServal 1468
4126 ACNCCTTATTTNAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCG 4175
1469 ThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGlyLy 1484
4176 AACACGCGTCAATACCGCNGTATTGGCTCAGGATTTCGGCAAAACCCGCA 4225
1484 sIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnGlnG 1501
4226 GTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCNTC 4275
1501 lnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1517

```
     4276 CACGCTGCCGCCGCCAAAGGNCCGCAACTGGAAGCGCAACACAGCGCGGG 4325
          ::::::   :::|||:::   |||   |||   |||   :::|||
     1518 IleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGl 1534

     4326 CATCAAATTAGGCTAC 4341
          :::||||||::::::
     1534 uLeuLysLeuSerPhe 1539

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564

seq_documentation_block:
ID   AAY75564 standard; Protein; 1431 AA.
XX
AC   AAY75564;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 759 protein sequence SEQ ID NO:2602.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54326.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1234; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   1431 AA;

alignment_scores:
           Quality: 1522.00              Length:   1639
             Ratio:  1.770                Gaps:     51
 Percent Similarity:  52.471   Percent Identity: 27.639

alignment_block:
 US-09-303-518D-651 x AAY75564   ..

 Align seg 1/1  to: AAY75564  from: 1  to: 1431

      55 GGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGG 104
         |||   :::      ||||||||||||   :::|||      :::
      17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31

     105 CATTCTTCCCCAAGCTTGGGCGGGACACACTTATTTCGGCATCAACTACC 154
                                                  ::::::||||
      32 .....AspValAspTyrG 36

     155 AATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCG 204
         |||||:::||||||||||||||||||||||||   |||:::||||||||||
      36 lnTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52

     205 AAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAAT 254
         ::::::|||::::::   :::|||:::|||:::::::::|||:::   ::
      53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69

     255 GACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT...AACGGCG 301
         ::::      ||||||   ||||||   |||   :::|||   :::::::
      69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86

     302 TGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGC 351
         ::|||:::||||||      |||||||:::   ||||||   ||||||
      86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102

     352 GGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCA 401
         ||||||:::::::::::|||||||   :::      ||||||:::::::
     103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGluGl 119

     402 CCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATT 451
         :   :::::::|||:::::::|||:::|||                  |||
     119 nAlaTyrThrTyrArgLeuValSerArg.....AsnP 130

     452 CACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTC 501
           ||||||   :::   |||   |||:::||||||||||:::|||:::|||
     130 roHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeuVal 145

     502 ACAGATGCAGAACCTGTCGAAATGACGAGT..... 531
         |||:::   :::|||:::   ::::::|||
     146 ThrGluIleSerProThrAlaLeuSerServalProLeuLeuGlyAsnGl 162

     532 .GACATGAGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTG 580
          :::   ::::::|||:::|||   |||   :::::::::|||      |
     162 yGlnProLysAlaAsnAlaTyrLeuAspThrAspArgProProTyrPheV 179

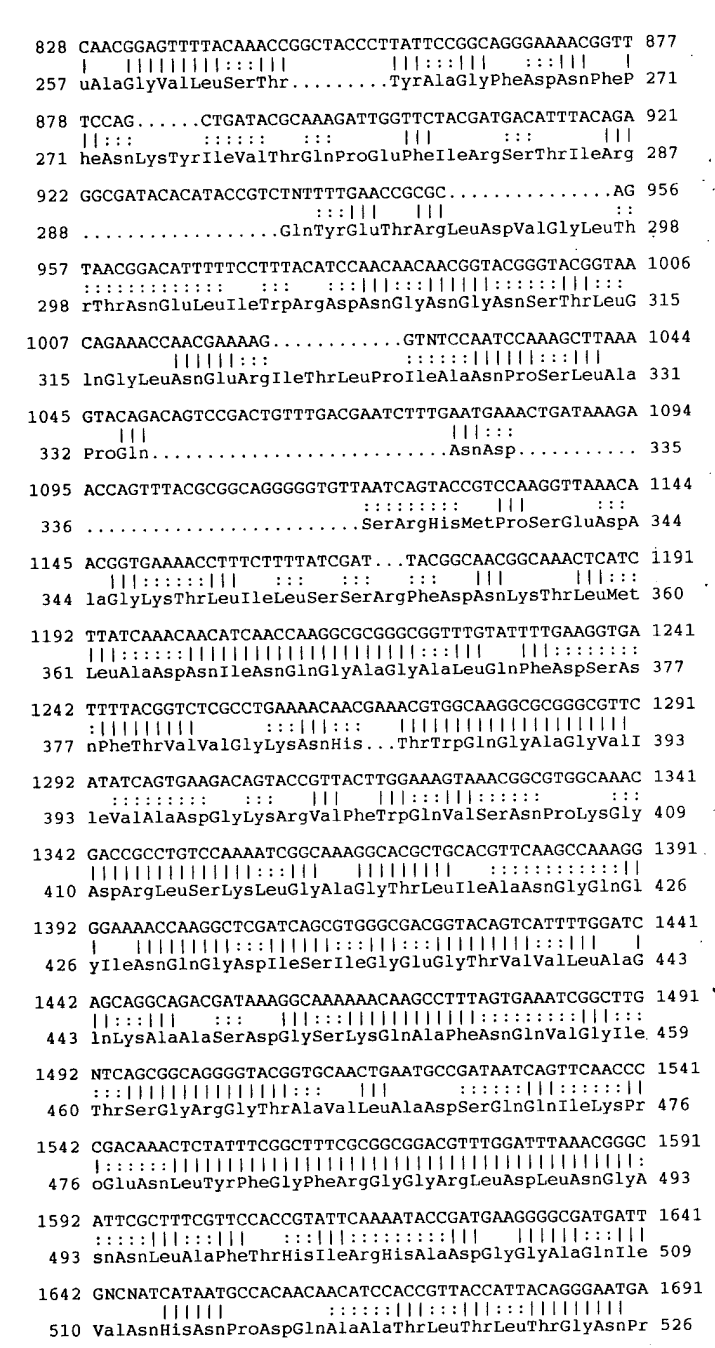
     581 TCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAACAC 630
         |||||:::|||||||||   :::      |||      |||      :::
     179 alArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGlyThrArg 195

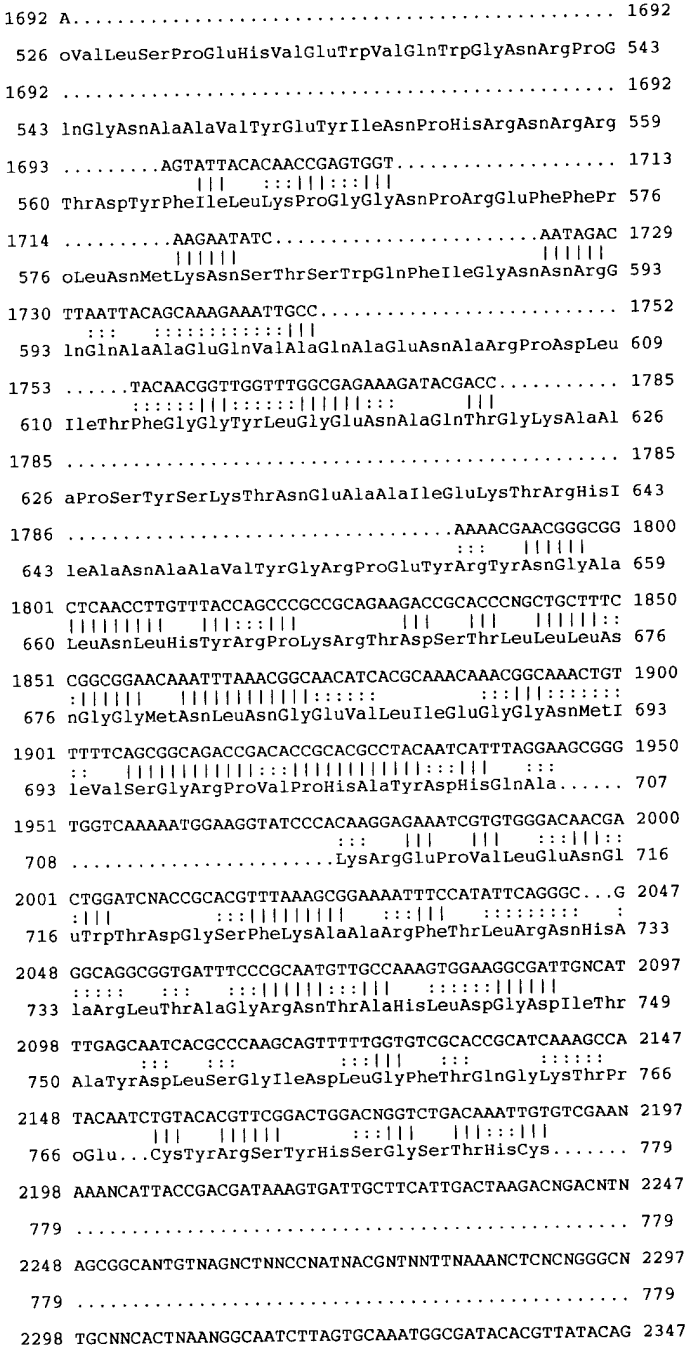
     631 GGCGATTTATCCTACTCCGGCGCATGGTTAATTGGCGGCAAT...ACACA 677
                :::   :::      :::|||   ||||||:::
     196 ThrArgThrAlaProAlaTyrGlnTyrLeuThrGlyGlyThrProLeuLy 212

     678 TATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGC 727
          :::   |||:::   |||:::|||:::      :::::::|||::::::
     212 sValLeuGlyPheGlnAsnHisGlyLeuLeu...ValGlyGlySerLeuT 228

     728 GCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGAC 777
                              |||:::          :::   |||||||||
     228 hrAspGln............ProLeuAsnThrTyrAlaIleAlaGlyAsp 240

     778 AGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCT 827
         ||||||||||||:::|||   :::|||||   :::|||:::|||:::||
     241 SerGlySerProLeuPheAlaPheAspLysHisGluAsnArgTrpValLe 257
```







```
3908 GCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTC 3957
     |||||::::::         ||||||::::::      :::
1286 rgArgHisAlaTrpAspAlaGlyIleAsnThrGlyIleLysIleAsp... 1301

3958 GGCGGATTCGGCATCGAACCGTACATCGGC..................... 3987
        |||:::::::::::::||||||   |||
1302 ThrGlyIleAsnLeuArgProTyrAlaGlyIleArgIleAsnArgSerAs 1318

3988 .GCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCA 4036
      ::::::||||||   ::::::   |||:::
1318 nGlyAsnArgTyrValLeuAspGlyAlaGlu..................... 1328

4037 ATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGCATTAAG 4086
       |||   :::|||:::         :::   ::::::||||||||||:::
1329 ..IleAsnSerProAlaGlnIleGlnThrThrTrpHisAlaGlyIleArg 1344

4087 GCAGATTATTCATTCAAACCGGCGCAACACATNTCCATCACNCCTTATTT 4136
       |||   :::   :::   :::|||   ::::::::::||||||
1345 LeuAspLysThrValGluLeuGlyGln...AlaLysLeuThrProAlaPh 1360

4137 NAGCCTGTCCTAT......ACCGATGCCGCTTCGGGCAAAGTCCGAACAC 4180
      |||   :::|||      |||         ||||||::::::   :::
1360 eSerSerAspTyrTyrHisThrArgGlnAsnSerGlySerAlaLeuSer. 1376

4181 GCGTCAATACCGCNGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCG 4230
       ||||||      :::|||   |||:::   :::
1377 ..ValAsnAspArgThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHis 1392

4231 GAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCNTCCACGC 4280
             :::::::|||      ||||||::::::   :::   |||||
1393 ThrLeuGlnIleAspAlaGlyTyrLysGlyTrpAsnAlaLysLeuHisAl 1409

4281 TGCCGCCGCCAAAGGNCCGCAACTGGAAGCGCAACACAGCGCGGGCATCA 4330
     ||||   :::|||      :::         :::   :::||||||||||
1409 aAlaTyrGlyLysAspSerAsnThrAlaArgHisLysGlnAlaGlyIleL 1426

4331 AATTAGGCTACCGCTGG 4347
     ||:::||||||:::|||
1426 ysIleGlyTyrAsnTrp 1431
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745

seq_documentation_block:

ID ABB52745 standard; Protein; 1371 AA.
XX
AC ABB52745;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 891.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP03445.
XX
PR 10-MAR-2000; 2000FR-0003145.
PR 02-FEB-2001; 2001FR-0001449.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the
PT phylogenic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
SQ Sequence 1371 AA;

alignment_scores:

Quality: 1056.50 Length: 1572
Ratio: 1.360 Gaps: 65
Percent Similarity: 49.427 Percent Identity: 26.272

alignment_block:

US-09-303-518D-651 x ABB52745 ..

Align seg 1/1 to: ABB52745 from: 1 to: 1371

```
 112 CCCCAA...GCTTGGGCGGGACACACTTATTTCGGCATCAACTACCAATA 158
     ||||||   :::   ||||||   :::         |||   ||||||
  50 ProGlnIleSerGlnAlaGlyIleValArgSerAspIleAlaTyrGlnIl 66

 159 CTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAG 208
     ||||||||||||||||||||||||||||   |||:::   ||||||   |
  66 eTyrArgAspPheAlaGluAsnLysGlyLeuPheValProGlyAlaThrA 83

 209 ATATTGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACA 258
     |||||   ||||||:::|||   |||:::||||||||||:::   :::
  83 spIleProValTyrAspLysAspGlyLysLeuValGlyArg...LeuAsp 98

 259 AAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGC 308
     ||||||||||||   |||||||||   ||||||   ||||||||||||::
  99 LysAlaProMetAlaAspPheSerSerValSerSerAsnGlyValAlaTh 115

 309 ATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATA 358
     :|||||:::   ||||||||||||||||||   ||||||||||||||:
 115 rLeuValSerProGlnTyrIleValSerValLysHisAsnGlyGlyTyrG 132

 359 ACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTT 408
     :::::|||:::||||||   :::|||   |||
 132 lnSerValSerPheGly...AsnGlyLysAsn..................... 141

 409 TCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCC 458
     :::|||:::::|||   |||||||||::::::
 142 ThrTyrSerLeuValAspArgAsnAsnHisSer..................... 152

 459 TTACAACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCACAGATG 508
        :::   |||   |||   |||||||||:::|||:::|||||||::::
 153 ....SerValAspPheHisAlaProArgLeuAsnLysLeuValThrGluV 168

 509 CAGAACCTGTCGAAATGACGAGTGAC...ATGAGGGGGAATACCTATTCC 555
     ::   |||      :::|||||||:::      ::::::|||:::|||:::
 168 alIleProSerAlaIleThrSerGluGlyThrLysAlaAsnAlaTyrLys 184

 556 GATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTA 605
     |||   |||:::|||            |||:::|||||||||   :::||
 185 AspThrGluArgTyrThrAlaPheTyrArgValGlySerGlyThrGlnTy 201
```





```
 3797 CCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGC 3846
      ::|||:::                  :::::::|||            |||
 1164 erHisAlaPheSerGlyLysThrLysSerValGlyGly.........Gly 1177

 3847 ATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGG 3896
      :::   :::::::|||   |||:::||||||:::      |||   |||||
 1178 LeuTyrAlaSerAlaLeuPheGluSerGlyAlaTyrIleAspLeuIleGl 1194

 3897 AGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACC 3946
      |                 :::   :::|||:::      :::      |||
 1194 y..............LysTyrIleHisHis......AspAsnAspTyrT 1204

 3947 GCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGC 3996
      ::::::|||:::|||:::|||   :::   |||         :::
 1204 hrGlyAsnPheAlaGlyLeuGlyThrLysHisTyrAsnThrHisSerTrp 1220

 3997 TATTTCGTCCAAAAAGCGGATTACCGCTAC..................... 4026
      |||        ::::::   |||||||||
 1221 TyrAlaGlyAlaGluThrGlyTyrArgTyrHisLeuThrGluGluThrPh 1237

 4027 ..............GAAAACGTCAATATCGCCACCCCCGGTCTTGCGT 4060
                     |||   |||      |||:::   |||   :::|
 1237 eIleGluProGlnAlaGluLeuValTyrGlyAlaValSerGlyLysThrP 1254

 4061 TCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCG 4110
      ||   |||:::::::   |||       |||   |||:::|||
 1254 he...ArgTrpLysAspGly...AspMetAspLeuSerMetLysAsnArg 1268

 4111 CAACACATNTCCATCACNCCTTATTTNAGCCTGTCCTATACCGATGCCGC 4160
      :::   :::   :::          ::::::::::::      :::
 1269 AspPheSerProLeuIleGlyArgThrGlyIleGluLeuGlyLysThrPh 1285

 4161 TTCGGGCAAA.........GTCCGAACACGCGTCAATACCGCN....... 4194
       |||||||||         |||   :::|||:::::::|||:::
 1285 eSerGlyLysAspTrpSerValThrAlaArgAlaGlyThrSerTrpGlnP 1302

 4195 .........................GTATTGGCTCAGGATTTCGGCAAA 4218
                               ||||||   :::      |||:::
 1302 heAspLeuLeuAsnAsnGlyGluThrValLeuArgAspAlaSerGlyGlu 1318

 4219 ACCCGCAGTGCG.........................GAATGGGGCGT 4241
        |||                                :::   |||::
 1319 LysArgIleLysGlyGluLysAspSerArgMetLeuPheAsnValGlyMe 1335

 4242 AAACGCCGAAATCAAA 4257
      :||||||:::||||||
 1335 tAsnAlaGlnIleLys 1340
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592

seq_documentation_block:

```
ID   ABB52592 standard; Protein; 1376 AA.
XX
AC   ABB52592;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 560.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1376 AA;
```

alignment_scores:

```
            Quality: 1038.00              Length:   1537
              Ratio:  1.294                 Gaps:     61
 Percent Similarity:  52.180   Percent Identity: 25.634
```

alignment_block:

US-09-303-518D-651 x ABB52592 ..

Align seg 1/1 to: ABB52592 from: 1 to: 1376

```
   10 ACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAA......AC 53
      |||   |||   :::::::|||      |||::::::   |||      ||
   15 ThrValLysValValSerGluLeuAlaArgArgValCysLysGlySerTh 31

   54 CGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCG 103
      |   |||   :::          :::      :::::::::||||||   :
   31 rArgArgGlyLysArgLeuSerValLeuThrSerLeuAlaLeuSer...A 47

  104 GCATTCTTCCCCAAGCTTGGGCGGGACACACTTATTTCGGCATCAAC... 150
      :::::||||||   :::   ::::::   |||      |||   |||
   47 laLeuLeuProThrValAlaGlyAlaSerThrValGlyGlyAsnAsnPro 63

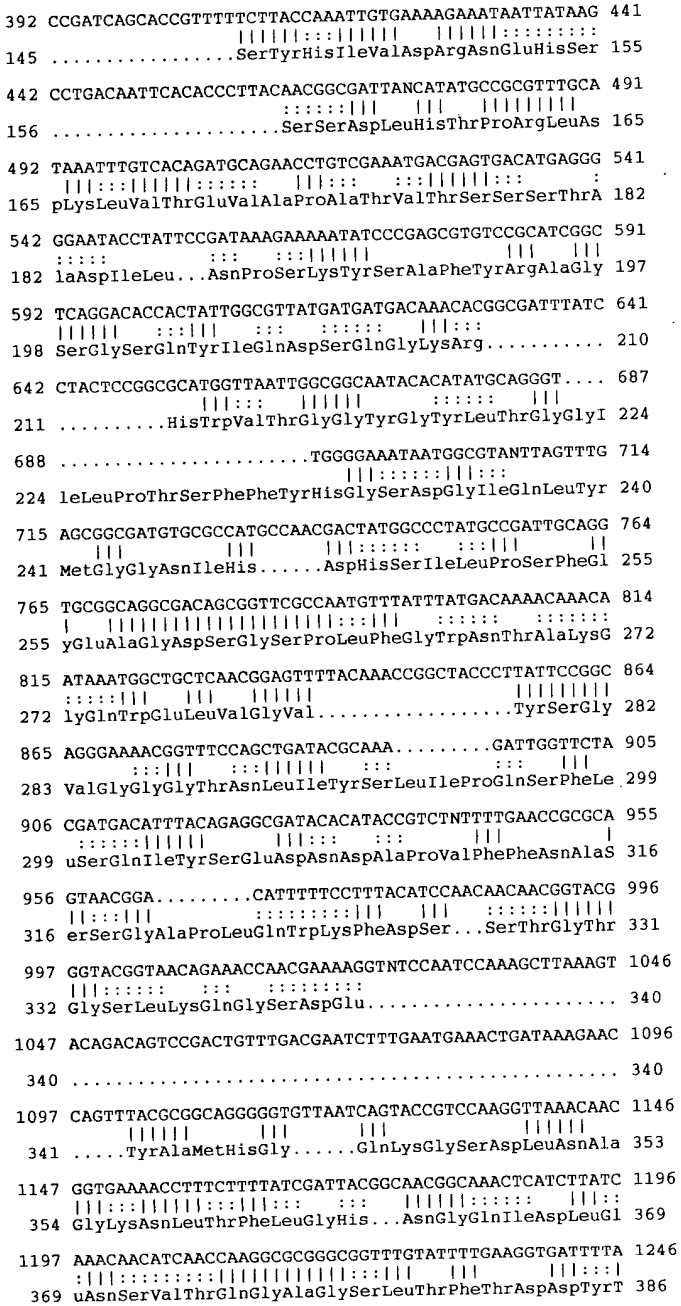
  151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
      ||||||   ||||||||||||||||||||||||||:::|||   :::||
   64 TyrGlnThrTyrArgAspPheAlaGluAsnLysGlyGlnPheGlnAlaGl 80

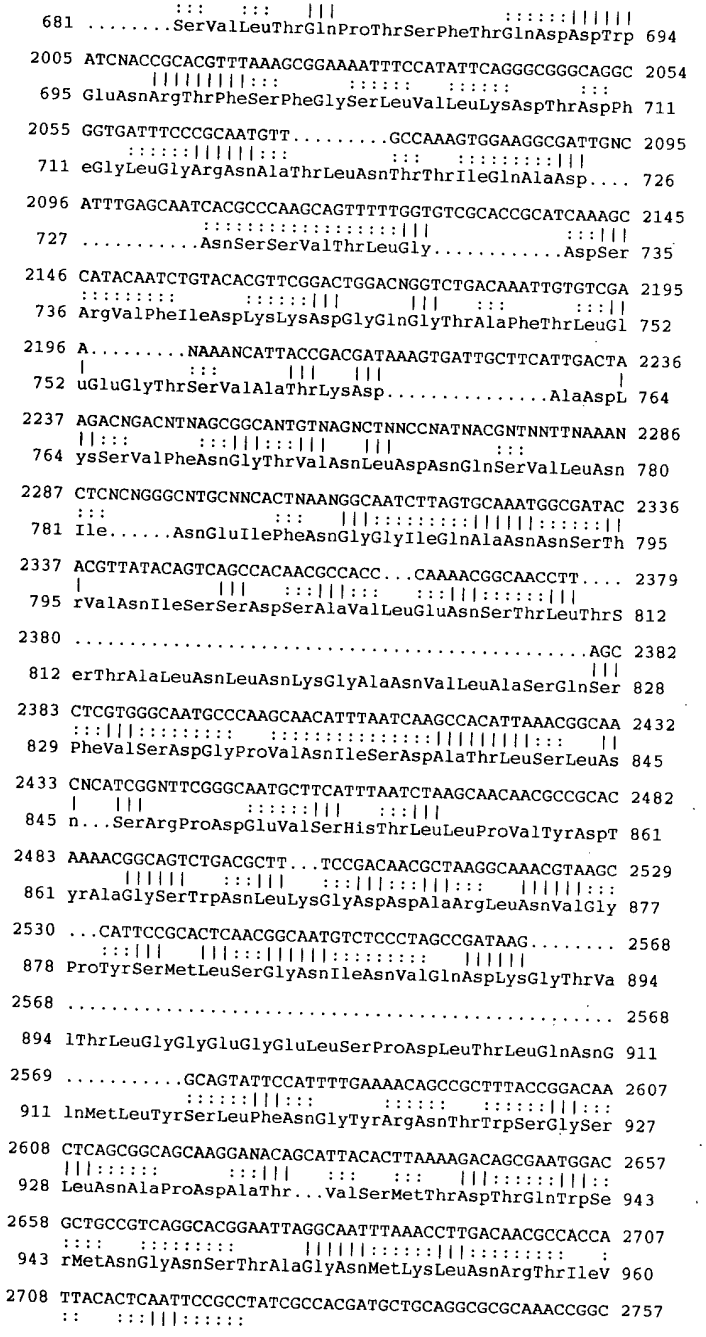
  201 GGCGAAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
      ||||   :::|||   ::::::|||:::|||||||||||||||||||
   80 yAlaThrAsnIleProIlePheAsnAsnLysGlyGluLeuValGly...H 96

  251 CAATGACAAAAGCCCCGATGATTGATTTTTCT.........GTGGTGTCG 291
        :::   ||||||||||||::|||||||||          |||   |||
   96 isLeuAspLysAlaProMetValAspPheSerSerValAsnValSerSer 112

  292 CGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGC 341
      :::   |||||||||:::|||:::::   |||||||||:::||||||
  113 AsnProGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSerValLy 129

  342 ACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGNAATC 391
      ||||||   ||||||:::::|||:::||||||   :::|||   |||
  129 sHisAsnLysGlyTyrGlnSerValSerPheGly...AspGlyGlnAsn. 144
```


392 CCGATCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAG 441
145 SerTyrHisIleValAspArgAsnGluHisSer 155
442 CCTGACAATTCACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCA 491
156 SerSerAspLeuHisThrProArgLeuAs 165
492 TAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGGG 541
165 pLysLeuValThrGluValAlaProAlaThrValThrSerSerSerThrA 182
542 GGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGC 591
182 laAspIleLeu...AsnProSerLysTyrSerAlaPheTyrArgAlaGly 197
592 TCAGGACACCACTATTGGCGTTATGATGATGACAAACACGGCGATTTATC 641
198 SerGlySerGlnTyrIleGlnAspSerGlnGlyLysArg..... 210
642 CTACTCCGGCGCATGGTTAATTGGCGGCAATACACATATGCAGGGT.... 687
211 HisTrpValThrGlyGlyTyrGlyTyrLeuThrGlyGlyI 224
688 TGGGGAAATAATGGCGTANTTAGTTTG 714
224 leLeuProThrSerPhePheTyrHisGlySerAspGlyIleGlnLeuTyr 240
715 AGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGG 764
241 MetGlyGlyAsnIleHis.....AspHisSerIleLeuProSerPheGl 255
765 TGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACA 814
255 yGluAlaGlyAspSerGlySerProLeuPheGlyTrpAsnThrAlaLysG 272
815 ATAAATGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGC 864
272 lyGlnTrpGluLeuValGlyVal.....TyrSerGly 282
865 AGGGAAAACGGTTTCCAGCTGATACGCAAA.....GATTGGTTCTA 905
283 ValGlyGlyGlyThrAsnLeuIleTyrSerLeuIleProGlnSerPheLe 299
906 CGATGACATTTACAGAGGCGATACACATACCGTCTNTTTTGAACCGCGCA 955
299 uSerGlnIleTyrSerGluAspAsnAspAlaProValPhePheAsnAlaS 316
956 GTAACGGA.....CATTTTTCCTTTACATCCAACAACAACGGTACG 996
316 erSerGlyAlaProLeuGlnTrpLysPheAspSer...SerThrGlyThr 331
997 GGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGT 1046
332 GlySerLeuLysGlnGlySerAspGlu..... 340
1047 ACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAAC 1096
340 340
1097 CAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGGTTAAACAAC 1146
341 TyrAlaMetHisGly.....GlnLysGlySerAspLeuAsnAla 353
1147 GGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATC 1196
354 GlyLysAsnLeuThrPheLeuGlyHis...AsnGlyGlnIleAspLeuGl 369
1197 AAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTA 1246
369 uAsnSerValThrGlnGlyAlaGlySerLeuThrPheThrAspAspTyrT 386
1247 CGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATC 1296
386 hrValThrThrSerAsnGlySerThrTrpThrGlyAlaGlyIleIleVal 402
1297 AGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1346
403 AspLysAspAlaSerValAsnTrpGlnValAsnGlyValLysGlyAspAs 419
1347 CCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAA 1396
419 nLeuHisLysIleGlyGluGlyThrLeuValValGlnGlyThrGlyValA 436
1397 ACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAG 1446
436 snGluGlyGlyLeuLysValGlyAspGlyThrValValLeuAsnGlnGln 452
1447 GCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAG 1496
453 AlaAspSerSerGlyHisValGlnAlaPheSerSerValAsnIleAlaSe 469
1497 CGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACA 1546
469 rGlyArgProThrValValLeuAlaAspAsnGlnGlnValAsnProAspA 486
1547 AACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCG 1596
486 snIleSerTrpGlyTyrArgGlyGlyValLeuAspValAsnGlyAsnAsp 502
1597 CTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGNCNA 1646
503 LeuThrPheHisLysLeuAsnAlaAlaAspTyrGlyAlaThrLeu...Gl 518
1647 TCATAATGCCACAACAACATCCACCGTTACCATTACAGGGAATGAAAGTA 1696
518 yAsnSerSerAspLysThrAlaAsnIleThrLeu.....AspTyrG 532
1697 TTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTACAGCAAAGAA 1746
532 lnThrArgProAlaAspValLysValAsnGluTrpSerSerSerAsnArg 548
1747 ATTGCCTACAAC..... 1758
549 GlyThrValGlySerLeuTyrIleTyrAsnAsnProTyrThrHisThrVa 565
1759 GGTTGGTTTGGCGAGAAAG 1777
565 lAspTyrPheIleLeuLysThrSerSerTyrGlyTrpPheProThrGlyG 582
1778 ATACGACCAAAACG..... 1791
582 lnValSerAsnGluHisTrpGluTyrValGlyHisAspGlnAsnSerAla 598
1792 AACGG 1796
599 GlnAlaLeuLeuAlaAsnArgIleAsnAsnLysGlyTyrLeuTyrHisGl 615
1797 GCGGCTC......AACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCC 1840
615 yLysLeuLeuGlyAsnIleAsnPheSerAsnLysAlaThrProGlyThrT 632
1841 NG.....CTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACG 1881
632 hrGlyAlaLeuValMetAspGlySerAlaAsnMetSerGlyThrPheThr 648
1882 CAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCC.. 1929
649 GlnGluAsnGlyArgLeuThrIleGlnGlyHisProValIleHisAlaSe 665
1930 TACAATCATTTAGGAAGCGGGTGGT 1954
665 rThrSerGlnSerIleAlaAsnThrValSerSerLeuGlyAspAsn.... 680
1955 CAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGG 2004


681 SerValLeuThrGlnProThrSerPheThrGlnAspAspTrp 694
2005 ATCNACCGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGC 2054
695 GluAsnArgThrPheSerPheGlySerLeuValLeuLysAspThrAspPh 711
2055 GGTGATTTCCCGCAATGTT.....GCCAAAGTGGAAGGCGATTGNC 2095
711 eGlyLeuGlyArgAsnAlaThrLeuAsnThrThrIleGlnAlaAsp.... 726
2096 ATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGC 2145
727 AsnSerSerValThrLeuGly.....AspSer 735
2146 CATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGTGTCGA 2195
736 ArgValPheIleAspLysLysAspGlyGlnGlyThrAlaPheThrLeuGl 752
2196 A.....NAAANCATTACCGACGATAAAGTGATTGCTTCATTGACTA 2236
752 uGluGlyThrSerValAlaThrLysAsp.....AlaAspL 764
2237 AGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAAAN 2286
764 ysSerValPheAsnGlyThrValAsnLeuAspAsnGlnSerValLeuAsn 780
2287 CTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCAAATGGCGATAC 2336
781 Ile.....AsnGluIlePheAsnGlyGlyIleGlnAlaAsnAsnSerTh 795
2337 ACGTTATACAGTCAGCCACAACGCCACC...CAAAACGGCAACCTT.... 2379
795 rValAsnIleSerSerAspSerAlaValLeuGluAsnSerThrLeuThrS 812
2380 AGC 2382
812 erThrAlaLeuAsnLeuAsnLysGlyAlaAsnValLeuAlaSerGlnSer 828
2383 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2432
829 PheValSerAspGlyProValAsnIleSerAspAlaThrLeuSerLeuAs 845
2433 CNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCAC 2482
845 n...SerArgProAspGluValSerHisThrLeuLeuProValTyrAspT 861
2483 AAAACGGCAGTCTGACGCTT...TCCGACAACGCTAAGGCAAACGTAAGC 2529
861 yrAlaGlySerTrpAsnLeuLysGlyAspAspAlaArgLeuAsnValGly 877
2530 ...CATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAG..... 2568
878 ProTyrSerMetLeuSerGlyAsnIleAsnValGlnAspLysGlyThrVa 894
2568 2568
894 lThrLeuGlyGlyGluGlyGluLeuSerProAspLeuThrLeuGlnAsnG 911
2569 GCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAA 2607
911 lnMetLeuTyrSerLeuPheAsnGlyTyrArgAsnThrTrpSerGlySer 927
2608 CTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGAC 2657
928 LeuAsnAlaProAspAlaThr...ValSerMetThrAspThrGlnTrpSe 943
2658 GCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCA 2707
943 rMetAsnGlyAsnSerThrAlaGlyAsnMetLysLeuAsnArgThrIleV 960
2708 TTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGC 2757
960 alGlyPheAsnGlyGly..... 965
2758 AGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGT 2807
965 965
2808 TACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACG 2857
966 ThrSerSerPheThrThrLeuThrThrAspA 976
2858 GCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGC 2907
976 snLeuAspAlaValGlnSerAlaPheValMetArgThrAspLeu..... 990
2908 TACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACAC 2957
991 AsnLysAlaAspLysLeuValIleAsnLysSerAlaThrGlyHisAspAs 1007
2958 CTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGA 3007
1007 nSerIleTrpValAsnPheLeuLysLysProSerAspLysAspThrLeuA 1024
3008 CGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTT..... 3051
1024 spIleProLeuValSerAlaProGluAlaThrAlaAspAsnLeuPheArg 1040
3052 AATTTCACCCTGCAAAA 3068
1041 AlaSerThrArgValValGlyPheSerAspValThrProThrLeuServa 1057
3069 CGAACACGTCGATGCC...GGCGCGTGG.....CGTTACCAACTCA 3106
1057 lArgLysGluAspGlyLysLysGluTrpValLeuAspGlyTyrGlnValA 1074
3107 TCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAG 3156
1074 laArgAsnAspGlyGln..... 1079
3157 CTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGA 3206
1080 GlyLysAlaAlaAlaThrPheMetHisIleSerTy 1091
3207 CAAC.....GCGCAAAGCCTTGACGCGCTGATTGCGGCCG 3241
1091 rAsnAsnPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspL 1108
3242 GGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNG 3291
1108 euArgAspIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnGly 1124
3292 GCAGGCGGGGAAAATGTCGGC.....ATTATGCAGGC 3323
1125 SerGlySerAlaAspGlyGlyPheThrAspHisTyrThrLeuLeuGlnMe 1141
3324 GGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC..... 3360
1141 tGlyAlaAspArgLysHisGluLeuGlySerMetAspLeuPheThrGlyV 1158
3361 AGCGCNTTGGCGAAACAGCGCGAAGCG 3387
1158 alMetAlaThrTyrThrAspThrAspAlaSerAlaGlyLeuTyrSerGly 1174
3388 GAAACCCGGCCGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCG 3437
1175 LysThrLysSerTrpGlyGlyGlyPheTyrAlaSerGlyLeuPheArgSe 1191
3438 GGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGC 3487
1191 rGly.....AlaTyrP 1195
3488 GCGACCTGATNAGCCGTTAT..... 3507
1195 heAspLeuIleAlaLysTyrIleHisAsnGluAsnLysTyrAspLeuAsn 1211

```
  3508 ...GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTT 3554
         |||::::::|||   :::::::|||   :::       :::|||:::::
  1212 PheAlaGlyAlaGlyLysGlnAsnPheArgSer......HisSerLeuTy 1226

  3555 CGCCGTACAGGACGAA.................TTGGACCGCGTGTTTG 3586
       :|||      :::                    |||   :::|||:
  1226 rAlaGlyAlaGluValGlyTyrArgTyrHisLeuThrAspThrThrPheV 1243

  3587 CCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGNACACCAAA 3636
       ::|||   :::   :::   ||||||                   |||
  1243 alGluProGlnAlaGluLeuValTrpGlyArgLeuGlnGlyGlnThrPhe 1259

  3637 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3686
       ::::::::::::::   :::
  1260 AsnTrpAsnAspSerGlyMet........................... 1266

  3687 CCAAATCGGTATGCAGAAAAACCTCGGCAGC.........GGGCGCGTCG 3727
        :::::::::|||::::::|||      :::          |||||||:::|
  1267 .AspValSerMetArgArgAsnSerValAsnProLeuValGlyArgThrG 1283

  3728 GCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCATC 3777
       ||:::::::         :::::::|||   :::      |||   ::::::
  1283 lyValVal......SerGlyLysThrPheSerGlyLysAspTrpSerLeu 1297

  3778 GGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCAT 3827
                |||   |||   :::
  1298 ThrAlaArgAlaGlyLeuHisTyr........................... 1305

  3828 CGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCA 3877
          :::|||||||:::      |||:::   :::   :::::::::::::
  1306 ....GluPheAspLeuThrAspSerAlaAspValHisLeuLysAspAlaA 1321

  3878 NTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCCGCGTGCTGCAT... 3924
       ::   :::      |||:::|||:::         |||:::|||:::
  1321 laGlyGluHisGlnIleAsnGlyArgLysAspGlyArgMetLeuTyrGly 1337

  3925 TACGGCATTCAGGCACGATAC.....................          3945
          |||::::::||||||:::
  1338 ValGlyLeuAsnAlaArgPheGlyAspAsnThrArgLeuGlyLeuGluVa 1354

  3946 ....CGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAA 3991
           |||::::::|||||   ::::::   :::      |||:::|||:
  1354 lGluArgSerAlaPheGlyLysTyrAsnThrAspAspAlaIleAsnAlaA 1371

  3992 CG...CGCTAT 3999
       ::   ||||||
  1371 snIleArgTyr 1374
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705

seq_documentation_block:

ID AAW27705 standard; Protein; 323 AA.
XX
AC AAW27705;
XX
DT 08-MAY-1998 (first entry)
XX
DE H. influenzae Hap protein autotransporter membrane integration region.
XX
KW Hap protein; autotransporter; Gram-negative bacteria; diagnostic;
KW therapy; surface presented polypeptide.
XX
OS Haemophilus influenzae.
XX
PN WO9735022-A1.
XX
PD 25-SEP-1997.
XX
PF 15-MAR-1996; 96WO-EP01130.
XX
PR 15-MAR-1996; 96WO-EP01130.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Jose J, Maurer J, Meyer TF;
XX
DR WPI; 1997-480227/44.
DR N-PSDB; AAT88142.
XX
PT Presentation of peptide(s) on surface of Gram-negative bacteria -
PT via transformation with vector encoding signal peptide, presented
PT peptide and transporter domain of auto-transporter, producing
PT peptide libraries for epitope mapping
XX
PS Claim 8; Fig 9; 84pp; German.
XX
CC This sequence represents the H. influenzae Hap autotransporter membrane
CC integration region. This region is involved in a novel method which
CC allows the presentation of stable fusion polypeptides on the surface of
CC Gram-negative bacteria which can be released into the surrounding media.
CC The method can be used to produce a variegated population of
CC surface-presented polypeptides, so that bacteria expressing polypeptides
CC with particular properties can be identified and simultaneously selected,
CC e.g. for epitope mapping or selection of ligands with the highest
CC affinity for antibodies, major histocompatibility complex (MHC) molecules
CC or other components of the immune system. Selected polypeptides can be
CC used diagnostically, e.g. to screen sera or antibody banks, and
CC polypeptide expressing cells may be used as live vaccines. They may also
CC be used therapeutically, e.g. when the polypeptide is an antibody, to
CC remove or concentrate pollutants, inactivate toxins, prepare and process
CC food, prepare washing compositions and label cells. Selected bacteria can
CC be stored, reproduced and replicated on a large scale as individual
CC clones.
XX
SQ Sequence 323 AA;

alignment_scores:

Quality: 709.50 Length: 311
Ratio: 2.920 Gaps: 3
Percent Similarity: 78.135 Percent Identity: 44.051

alignment_block:

US-09-303-518D-651 x AAW27705 ..

Align seg 1/1 to: AAW27705 from: 1 to: 323

```
  3439 GATTTGCCGCAA.....................CCGCAGCCCCAACCGCA 3467
       |||   ||||||                     |||   :::   ::
    14 AspAlaProGlnGlnSerGluLysAspArgLeuAlaGlnGluGluAlaGl 30

  3468 ACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGCG 3517
       :   |||   :::   |||:::||||||   |||||||||:::|||||||:
    30 uLysGln...ArgLysGlnLysAspLeuIleSerArgTyrSerAsnSerA 46

  3518 GTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGAC 3567
       ::|||||||||:::|||||||||:::||||||::::::::::|||||||||
    46 laLeuSerGluLeuSerAlaThrValAsnSerMetLeuSerValGlnAsp 62

  3568 GAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAG 3617
       ||||||||||||:::|||:::::::::   ::::::||||||||||||::
    63 GluLeuAspArgLeuPheValAspGlnAlaGlnSerAlaValTrpThrAs 79

  3618 CNGCATCCGGNACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACC 3667
       :      :::      :::::::|||   |||:::   ||||||||||||:
    79 nIleAlaGlnAspLysArgArgTyrAspSerAspAlaPheArgAlaTyrG 96

  3668 GCCAACAA...ACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3714
       ::||||||   |||:::|||||||||||||||:::||||||   |||:::
    96 lnGlnGlnLysThrAsnLeuArgGlnIleGlyValGlnLysAlaLeuAla 112

  3715 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTT 3764
       :::|||||||:::|||   :::|||||||||:::|||:::::::|||:::||
   113 AsnGlyArgIleGlyAlaValPheSerHisSerArgSerAspAsnThrPh 129
```

```
 3765 CGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCG 3814
      ||||:::  :::  |||  |||  |||:::     :::  |||:
  129 eAspGluGlnValLysAsnHisAlaThrLeuThrMetMetSerGlyPheA 146

 3815 GGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT 3864
      ::||||||     |||  :::::::::|||:::::::::|||:::|||
  146 laGlnTyrGlnTrpGlyAspLeuGlnPheGlyValAsnValGlyThrGly 162

 3865 TTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCG 3914
      :::|||:::::  :::::::::      :::  ||||||:::|||::
  163 IleSerAlaSerLysMetAlaGluGluGlnSerArgLysIleHisArgLy 179

 3915 CGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGAT 3964
      ::::::::::||||||::::::|||   |||:::     :::|||  :
  179 sAlaIleAsnTyrGlyValAsnAlaSerTyrGlnPheArgLeuGlyGlnL 196

 3965 TCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4014
      ::||||||:::||||||:::|||:::::::||||||||:::::::::
  196 euGlyIleGlnProTyrPheGlyValAsnArgTyrPheIleGluArgGlu 212

 4015 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAA 4064
      :::|||:::   |||:::|||:::::   ||||||:::||||||||||
  213 AsnTyrGlnSerGluGluValArgValLysThrProSerLeuAlaPheAs 229

 4065 CCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAAC 4114
      |||||||:::|||||||||:::::::||||||:::|||   |||:::::::
  229 nArgTyrAsnAlaGlyIleArgValAspTyrThrPheThrProThrAspA 246

 4115 ACATNTCCATCACNCCTTATTTNAGCCTGTCCTATACCGATGCCGCTTCG 4164
      ::   |||:::   ||||||      ::::::|||:::|||:::::::::
  246 snIleSerValLysProTyrPhePheValAsnTyrValAspValSerAsn 262

 4165 GGCAAAGTCCGAACACGCGTCAATACCGCNGTATTGGCTCAGGATTTCGG 4214
      :::::::|||:::|||   ||||||   :::||||||   |||   |||||
  263 AlaAsnValGlnThrThrValAsnLeuThrValLeuGlnGlnProPheGl 279

 4215 CAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCA 4264
      |:::      :::   |||   |||::::::|||||||||      |||
  279 yArgTyrTrpGlnLysGluValGlyLeuLysAlaGluIleLeuHisPheG 296

 4265 CGCTGTCCNTCCACGCTGCCGCCGCCAAAGGNCCGCAACTGGAAGCGCAA 4314
        :::|||:::      :::   ::::::|||   ||||||      |||
  296 lnIleSerAlaPheIleSerLysSerGlnGlySerGlnLeuGlyLysGln 312

 4315 CACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4347
      :::::::::|||:::|||||||||||||||||
  313 GlnAsnValGlyValLysLeuGlyTyrArgTrp 323

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136

seq_documentation_block:
ID   AAP80136 standard; protein;  741 AA.
XX
AC   AAP80136;
XX
DT   09-OCT-1990  (first entry)
XX
DE   Neisseria IgA-Protease precursor protein.
XX
KW   IgA-Protease precursor; Neisseria sp;
KW   Gram-negative bacterial live vaccines.
XX
OS   Neisseria gonorrhoeae (strain SMH).
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   193..197
FT                   /label=IgA-Protease cleavage site (a)
FT   Cleavage-site   226..229
FT                   /label=IgA-Protease cleavage site (b)
FT   Cleavage-site   327..332
FT                   /label=IgA-Protease cleavage site (c)
XX
PN   DE3622221-A.
XX
PD   14-JAN-1988.
XX
PF   02-JUL-1986;   86DE-3622221.
XX
PR   02-JUL-1986;   86DE-3622221.
XX
PA   (PLAC ) MAX PLANCK GES WISSENSCH.
XX
PI   Meyer TF,  Halter R,  Pohlner J;
XX
DR   WPI; 1988-015104/03.
DR   N-PSDB; AAN80154.
XX
PT   Extracellular prodn. of proteins -
PT   by gram-negative host cells contg. a vector contg. one or more
PT   genes coding the desired protein
XX
PS   Disclosure; ; ; German.
XX
CC   Precursor protein consists of three regions i.e. Amino terminal
CC   leader sequence, IgA protease and a "helper" domain. The cleavage
CC   sites given in the features lie in the region between the latter
CC   two domains. DNA encoding a desired protein can be cloned into the
CC   corresponding region in the IgA protease precursor gene, between
CC   the DNA that encodes the natural cleavage sites. Thus, IgA protease
CC   coding region is not disrupted and the desired protein is released
CC   following cleavage by the protease.
XX
SQ   Sequence   741 AA;


alignment_scores:
            Quality:  385.50              Length:    785
              Ratio:  1.083                 Gaps:     18
 Percent Similarity:  45.350   Percent Identity: 20.892

alignment_block:
 US-09-303-518D-651 x AAP80136   ..

 Align seg 1/1  to: AAP80136  from: 1  to: 741

 2500 CTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAA 2549
      ||||||||||:::|||      :::   :::   :::   :::||||||||
    1 LeuSerAspLysAlaLeuAsnSerPheAspAlaThrProIleAsnGlyAs 17

 2550 TGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTA 2599
      ||||:::|||   :::::::|||::::::   :::   ::::::   :::
   17 nValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGlyLysAlaAlaLeuT 34

 2600 CCGGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTA...AAAGAC 2646
        |||::::::::::|||:::      :::   :::   |||   :::
   34 rpGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis 50

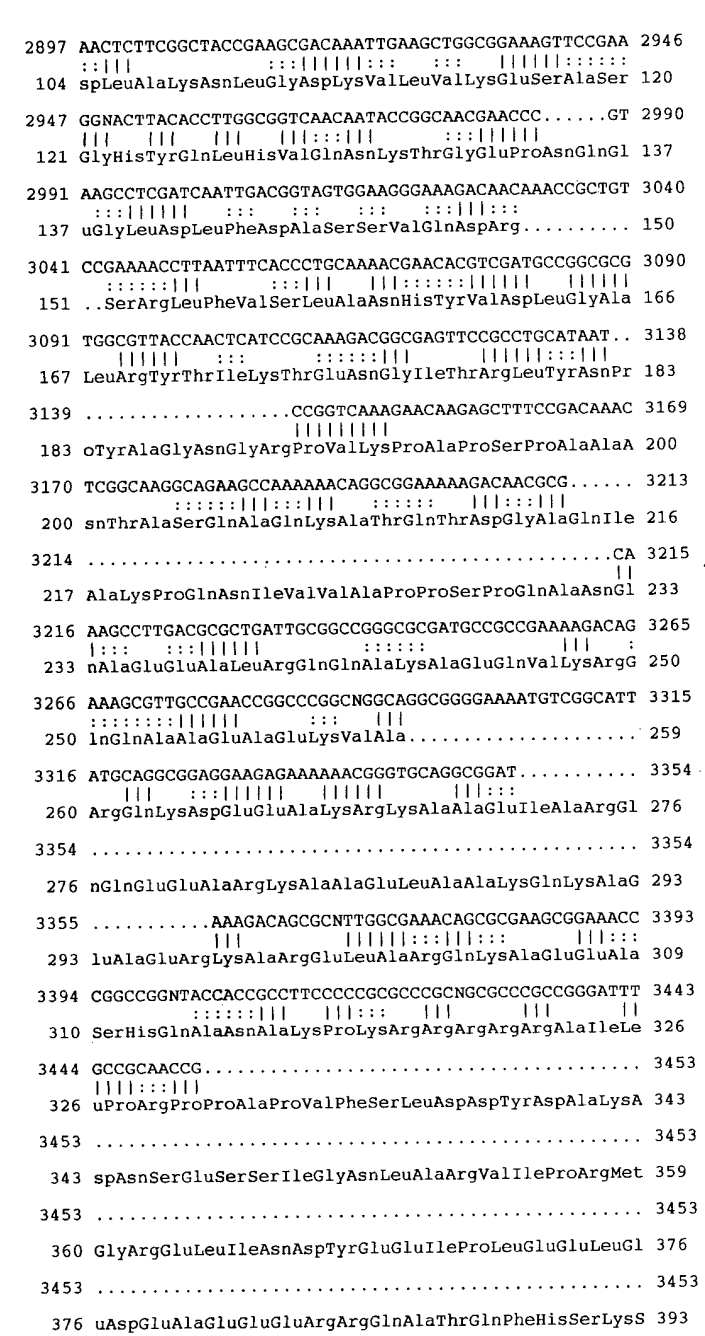
 2647 AGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGA 2696
      |||:::|||   |||   :::   :::::::::   ||||||:::|||
   51 SerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAl 67

 2697 CAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCG 2746
       ::::::   |||   ||||||:::|||          ||||||   :::|
   67 aAspSerHisIleHisLeuAsnAsnAla......SerAspAlaGlnSerA 82

 2747 CGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCC 2796
      ||
   82 la..................................................  82

 2797 CTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCT 2846
                                          :::::::::::|||::
   83 ..............................AsnLysTyrHisThrIl 88

 2847 GACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGG 2896
      :   :::|||      |||:::   :::|||   |||:::::::::::::
   88 eLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrA 104
```


2897 AACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAA 2946
104 spLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluSerAlaSer 120
2947 GGNACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCC.....GT 2990
121 GlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluProAsnGlnGl 137
2991 AAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGT 3040
137 uGlyLeuAspLeuPheAspAlaSerSerValGlnAspArg..... 150
3041 CCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3090
151 ..SerArgLeuPheValSerLeuAlaAsnHisTyrValAspLeuGlyAla 166
3091 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAAT.. 3138
167 LeuArgTyrThrIleLysThrGluAsnGlyIleThrArgLeuTyrAsnPr 183
3139 CCGGTCAAAGAACAAGAGCTTTCCGACAAAC 3169
183 oTyrAlaGlyAsnGlyArgProValLysProAlaProSerProAlaAlaA 200
3170 TCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCG..... 3213
200 snThrAlaSerGlnAlaGlnLysAlaThrGlnThrAspGlyAlaGlnIle 216
3214 CA 3215
217 AlaLysProGlnAsnIleValValAlaProProSerProGlnAlaAsnGl 233
3216 AAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAG 3265
233 nAlaGluGluAlaLeuArgGlnGlnAlaLysAlaGluGlnValLysArgG 250
3266 AAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATT 3315
250 lnGlnAlaAlaGluAlaGluLysValAla..... 259
3316 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGAT..... 3354
260 ArgGlnLysAspGluGluAlaLysArgLysAlaAlaGluIleAlaArgGl 276
3354 3354
276 nGlnGluGluAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAlaG 293
3355 AAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACC 3393
293 luAlaGluArgLysAlaArgGluLeuAlaArgGlnLysAlaGluGluAla 309
3394 CGGCCGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTT 3443
310 SerHisGlnAlaAsnAlaLysProLysArgArgArgArgArgAlaIleLe 326
3444 GCCGCAACCG..... 3453
326 uProArgProProAlaProValPheSerLeuAspAspTyrAspAlaLysA 343
3453 3453
343 spAsnSerGluSerSerIleGlyAsnLeuAlaArgValIleProArgMet 359
3453 3453
360 GlyArgGluLeuIleAsnAspTyrGluGluIleProLeuGluGluLeuGl 376
3453 3453
376 uAspGluAlaGluGluGluArgArgGlnAlaThrGlnPheHisSerLysS 393
3453 3453
393 erArgAsnArgArgAlaIleSerSerGluProSerSerAspGluAspAla 409
3453 3453
410 SerGluSerValSerThrSerAspLysHisProGlnAspAsnThrGluLe 426
3454 CAGCCC.....CAACCGC 3466
426 uHisGluLysValGluThrAlaGlyLeuGlnProArgAlaAlaGlnProA 443
3467 AACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGC 3516
443 rgThrGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSer 459
3517 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT..... 3552
460 AlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGl 476
3553 TTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCC 3598
476 yAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaAspAlaGluL 493
3599 GCAACGCNGTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCG 3648
493 ysAsnSerValTrpMetSerAsnThrGlyTyrGlyArgAspTyrAlaSer 509
3649 CAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTAT 3698
510 AlaGlnTyrArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIl 526
3699 GCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACA 3745
526 eAspArgSerLeuSerGluAsnMetGlnIleGlyGlyValLeuThrTyrS 543
3746 ACCGGACCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGCTT 3795
543 erAspSerGlnHisThrPheAspGlnAla...GlyGlyLysAsnThrPhe 558
3796 GCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGC...AGGTTCGACAT 3842
559 ValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrVa 575
3843 CGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCA 3892
575 lAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnG 592
3893 TCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGA 3942
592 lnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuThrLeuGly 608
3943 TACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAAC 3992
609 AsnThrLeuLysIleAsnGlnPheGluIleValProSerAlaGlyIleAr 625
3993 GCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG 4042
625 gTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAspAspSerValL 642
4043 CCACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGAT 4092
642 ysValSerSerMetAlaValLysThrLeuThrAlaGlyLeuAspPheAla 658
4093 TATTCATTCAAACCGGCGCAACACATNTCCATCACNCCTTATTTNAGCCT 4142
659 TyrArgPheLys...ValGlyAsnLeuThrValLysProLeuLeuSerAl 674
4143 GTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCG 4192
674 aAlaTyrPhe...AlaAsnTyrGlyLysGlyGlyValAsnValGlyGlyL 690
4193 CNGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTA 4242

```
          :::|||           :::   ::::::   :::   ||||||
 690 ysSerPheAlaTyrLysAlaAspAsnGlnGlnGlnTyrSerAlaGlyVal 706

4243 AACGCCGAAATCAAAGGTTTCACGCTGTCCNTCCACGCTGCCGCCGCCAA 4292
              ::::::   ||||||:::   :::::::::   :::||
 707 AlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLy 723

4293 AGGNCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACC 4342
     ||||   |||||||||   |||   |||:::   |||||||:::     |
 723 sGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleGlnIleA 740

4343 GCTGG 4347
     ||:::
 740 rgPhe 741
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842

seq_documentation_block:

ID AAG98842 standard; Protein; 1569 AA.
XX
AC AAG98842;
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related protein sequence SEQ ID NO:312.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
OS Escherichia coli.
XX
PN WO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zyskind J;
XX
DR WPI; 2001-335933/35.
DR N-PSDB; AAH84513.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors
PT -
XX
PS Claim 19; Page 396-399; 522pp; English.
XX
CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
CC growth and proliferation related proteins given in AAG99078 and AAG98830
CC to AAG98999. (I) can be used as potential targets for the generation of
CC new antimicrobial agents, and for identification of compounds which
CC interact with the gene products of (I). In addition the expression of
CC (I) and the purification of the proteins, the purified proteins can be
CC used to generate reagents and screen small molecule libraries or other
CC candidate compound libraries for compounds that can be further developed
CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (I) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable
CC method by which to identify the causative agents of a bacterial
CC infection. Also, antibodies generated against proteins translated from
CC mRNA transcribed from proliferation-required sequences can also be used
CC to screen for specific microorganisms that produce such proteins in a
CC species-specific manner. AAH84371 and AAH84670 represent sequencing
CC primers used in the isolation of E. coli growth and proliferation
CC related sequence, which are used in an example from the present
CC invention.

XX
SQ Sequence 1569 AA;

alignment_scores:
Quality: 372.50 Length: 1676
Ratio: 0.491 Gaps: 76
Percent Similarity: 45.286 Percent Identity: 19.153

alignment_block:
US-09-303-518D-651 x AAG98842 ..

Align seg 1/1 to: AAG98842 from: 1 to: 1569

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     |||   |||      :::   |||   :::   |||:::        ::
 160 GlyThrThrIleGluSerGlyAsnGlnAspValTyrLysGlyGlyIleSe 176

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
     :|||        :::      ||||||:::   :::|||      :::
 176 rAsnGlyThrThrIleLysGlyGlyAlaSerArgValGluGlyGlySerA 193

 227 AAAAAGGGGAGTTGGTC......GGCAAATCAATGACAAAAGCCCCG... 267
       :::|||   |||:::      |||:::::::::::|||:::
 193 laAsnGlyIleLeuIleAspGlyGlySerGlnIleValLysValGlnGly 209

 268 ATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTG.. 315
         |||   :::::::::::::::::::|||         :::|||
 210 HisAlaAspGlyThrThrIleAsnLysSerGlySerGlnAspValValGl 226

 316 ..............................GGCGATCAATATATT. 330
                                    |||   ||||||:::
 226 nGlySerLeuAlaThrAsnThrThrIleAsnGlyGlyArgGlnTyrValG 243

 331 ................GTGAGCGTGGCACATAACGGCGGCTATAACAAC 363
                     :::::::::      |||||||||   ::::::
 243 luGlnSerThrValGluThrThrThrIleLysAsnGlyGlyGluGlnArg 259

 364 GTT.................GATTTTGGTGCGGAAGGAAGNAATCCCGA 395
     |||                 |||         ||||||   :::   ::
 260 ValTyrGluSerArgAlaLeuAspThrThrIleGluGlyGlyThrGlnSe 276

 396 TCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTG 445
     :   :::      |||
 276 rLeuAsnSerLysSer.........................ThrAlaL 284

 446 ACAATTCACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAA 495
       |||:::|||   |||:::|||   :::         ::::::
 284 ysAsnThrHisIleTyrSerGlyGlyThr.........GlnIleValAsp 297

 496 TTTGTCACAGATGCAGAACCTGTCGAAATGACGAGT.............GA 533
     ::::::   ::::::   :::|||:::   |||             ||
 298 AsnThrSerThrSerAspValIleGluValTyrSerGlyGlyValLeuAs 314

 534 CATGAGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCC 583
     |:::|||||||:::|||   ::::::      :::
 314 pValArgGlyGlyThrAlaThrAsnValThrGln................ 325

 584 GCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAACACGGC 633

 325 ................................................. 325

 634 GATTTATCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACAT..... 678
               ::::::||||||   ::::::      |||||||::
 326 .........HisAspGlyAla...IleLeuLysThrAsnThrAsnGlyTh 338

 679 ....ATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTG.......... 714
         ::::::|||   :::::::::|||:::   |||:::
 338 rThrValSerGlyThrAsnSerGluGlyAlaPheSerIleHisAsnHisV 355

 715 ......................AGCGGCGATGTGCGCCATGCCAACGAC 741
                            :::|||               |||
 355 alAlaAspAsnValLeuLeuGluAsnGlyGlyHisLeuAspIleAsnAla 371
```

```
 742 TATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTTCGCCAAT 791
     ||||||                                 |||:::::
 372 TyrGly............................SerAlaAsnLysTh 378

 792 GTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGAGTTTTAC 841
     :::|||   ||||||   :::   :::   |||   |||:::      :
 378 rIleIleLysAspLysGlyThrMetSerValLeuThrAsnAlaLysAlaA 395

 842 AAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGC 891
     :::::           :::     :::|||||||      :::
 395 spAla............ThrArgIleAspAsnGlyGlyValMet..... 405

 892 AAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTCTN 941
                     |||:::     :::   |||:::|||:::
 406 ................AspValAlaGlyAsnAlaThrAsnThrIleIl 416

 942 TTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACG 991
                   ||||||     :::   :::   ||||||   |
 416 e.............AsnGlyGly...ThrGlnAsnIleAsnAsnTyrG 427

 992 GTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAGCTT 1041
     ||   :::|||   |||:::   |||:::   ::::::|||   |||
 427 lyIleAlaThrGlyThrAsnIleAsnSerGlyThrGlnAsnIleLysSer 443

1042 ......AAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAAC 1085
           |||::::::|||:::   :::   :::   |||   :::   ::
 444 GlyGlyLysAlaAspThrThrIleIleSerSerGlySerArgGlnValVa 460

1086 TGATAAAGAACCAGTT..............TACGCGGCAGGGGGTGTTA 1120
     :::::|||:::   :::                  :::|||||||||
 460 lGluLysAspGlyThrAlaIleGlySerAsnIleSerAlaGlyGly.... 475

1121 ATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGAT 1170
                                           |||:::|||
 476 ....................................SerLeuIleVal 479

1171 TACGGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGG 1220
     |||   :::|||         ::::::::::::::|||||   :::||
 480 TyrThrGlyGly.........IleAlaHisGlyValAsnGlnGluThrGl 493

1221 CGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAA 1270
     |:::          :::
 493 ySerAlaLeuValAlaAsn.............................. 499

1271 CGTGGCAAGGCGCGGGCGTTCATATCAGT..............GAAGAC 1305
            |||||||||:::   |||:::                :::
 500 .....ThrGlyAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHis 514

1306 AGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAA 1355
        |||:::|||         :::|||   ||||||
 515 PheThrIleThr.........GlyGlyGluAlaAsn.............. 523

1356 AATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCT 1405
                                    :::      ||||||   |||:
 524 ........................TyrValValLeuGluAsnThrGlyG 532

1406 CGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGAT 1455
     :::::::|||         |||          :::          |||
 532 luLeuThrValValAlaLysThrSerAlaLysAsnThrThrIleAspThr 548

1456 AAAGGCAAA............AAACAAGCCTTTAGTGAAATCGGCTTGNT 1493
        ||||||           |||:::|||   ::::::
 549 GlyGlyLysLeuIleValGlnLysGluAlaLysThrAsp........... 561

1494 CAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1543
                   :::::::::||||||
 562 ............SerThrArgLeuAsn..................... 566

1544 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1593
                         :::||||||   |||:::::::::
 567 ..................AsnGlyGlyValLeuGluValGlnAspGly 576

1594 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGN 1643
     :::   :::   |||   :::::::::   :::   ||||||:::|||
 577 GlyGluAlaLysHis...ValGluGlnGlnSerGlyGlyAlaLeuIle.. 591

1644 CNATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGAATGAAA 1693
              |||:::|||||||||   :::   |||   |||::::::|
 592 ..........AlaSerThrThrSerGlyThrLeuIleGluGlyThrAsnS 605

1694 GT.................ATTACACAACCGAGTGGTAAGAATATCAAT 1725
     ||                  |||   :::   ::::::||||||:::
 605 erTyrGlyAspAlaPheTyrIleArgAsnSerGluAlaLysAsnValVal 621

1726 AGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1775
           |||   ::::::     :::   :::|||      :::
 622 LeuGluAsnAlaGlySerLeuThrValValThrGlySerArgAlaValAs 638

1776 AGATACGACCAAAACGAACGGGCGGCTCAACCTT................ 1809
               ::::::||||||::::::::::::
 638 pThrIleIleAsnAlaAsnGlyLysMetAspValTyrGlyLysAspValG 655

1810 ..............................GTTTACCAGCCCGCCGCA 1827
                                    :::|||      |||:::
 655 lyThrValLeuAsnSerAlaGlyThrGlnThrIleTyrAlaSerAlaThr 671

1828 GAAGACCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACAT 1877
     :::|||::::::   :::   :::|||      :::::   |||
 672 SerAspLysAlaAsnIleLysGlyGlyLysGlnThrValTyrGlyLeuAl 688

1878 CACGCAAACAAAC.........GGCAAACTGTTTTTCAGCGGCAGACCGA 1918
      |||:::::::|||          |||:::   :::   :::|||     |
 688 aThrGluAlaAsnIleGluSerGlyGluGlnIleValAspGlyGlySerT 705

1919 CACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT 1968
     ||            :::|||::::::::::|||   :::   :::::::::
 705 hrGluLys......ThrHisIleAsnGlyGlyThrGlnThrValGlnAsn 719

1969 ATCCCACAAGGA.........GAAATCGTGTGGGACAACGACTGGATCNA 2009
           ::::::         :::||||||         :::   |||
 720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnGlnIleMe 736

2010 CCGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGA 2059
        :::      ||||||:::      |||:::||||||:::   |||:
 736 tAlaAsnGlyThrAlaGluGlySerIleIleAsnGlyGlySerGlnValV 753

2060 TTTCCCGC..................................... 2067
     ::::::::
 753 alAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnAspGlyGlyThr 769

2068 ...AATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCA 2114
        :::|||   :::      :::   :::   :::::::::   :::||
 770 LeuAspValArgGluLysGlySerAlaThrGlyIleGlnGlnSerSerGl 786

2115 AGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACACGTT 2164
     |:::::::::   ::::::   :::   :::::::::      |||||||:
 786 nGlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArgA 803

2165 CGGACTGGACN.................GGTCTGACAAAT......... 2187
     ::|||   :::                  |||   :::|||
 803 laAspGlyValAlaPheSerIleGluGlnGlyAlaAlaAsnAsnIleLeu 819

2188 .....................TGTGTCGAANAAANCATTACCGACGATAA 2216
                           |||||||:::      ::::::||||
 820 LeuAlaAsnGlyGlyValLeuThrValGluSerAspThrSerSerAspLy 836

2217 AGTGATTGCTTCATTG.....................ACTAAGACNGACN 2245
     |:::   :::::::::                    ||||||:::   :
 836 sThrGlnValAsnMetGlyGlyArgGluIleValLysThrLysAlaThrA 853

2246 TNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGG 2295
     :::::|||                        ::::::         |||
 853 laThrGly.......................ThrThrLeuThrGlyGly 861

2296 CNTGCNNCACTNAANGGC.............................. 2313
```


862 GluGlnIleValGluGlyValAlaAsnGluThrThrIleAsnAspGlyGl 878
2314 AATCTTAGTGCAAATGGCGAT.....ACACGTTATACAG 2347
878 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsnG 895
2348 TCAGCCACAACGCCACCCAAAACGGCAAC.....CTTAGCCTCGTG 2388
895 luGlyGlyThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleVal 911
2389 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACNCATC 2438
912 GlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly...IleGl 927
2439 GGNTTCGGGCAAT.....GCTTCATTTAATCTAAGCAACAACGCCG 2479
927 uIleSerGlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuA 944
2480 CACAAAAC.....GGCAGTCTGACGCTTTCCGACAAC 2511
944 laThrAsnMetLeuLeuGluAsnGlyGlyAsnLeuLeuValLeuAlaGly 960
2512 GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGC 2561
961 ThrGluAlaArgAspSerThrValGlyLysGlyGly..... 972
2562 CGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCA 2611
973 AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsnS 987
2612 GCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTG 2661
987 erGlyGlyGlnTyrThrLeuGlyArgSerLysAspGluPheGlnAlaLeu 1003
2662 CCGTCAGGCACGGAATTA..... 2679
1004 AlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTyrAl 1020
2680 GGCAATTTAAACC 2692
1020 aGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSerL 1037
2693 TT..... 2694
1037 euMetThrProArgAspAsnValThrProValLysLeuGluGlyAlaVal 1053
2695 GACAACGCCACCATTACACTCAATTCCGCCTAT..... 2727
1054 ArgIleThrAspSerAlaThrLeuThrLeuGlyAsnGlyValAspThrTh 1070
2728 .CGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTG..... 2763
1070 rLeuAlaAspLeuThrAlaAlaSerArgGlySerValTrpLeuAsnSerA 1087
2764 TCAGACACGCCGCGCCGCCGTTCGCGC...CGTTCCCTA 2799
1087 snAsnSerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSerLeu 1103
2800 TTA.....TCCGTTACACCGCCAAC 2819
1104 LeuLeuAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaAlaProAl 1120
2820 TTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACN 2869
1120 aThrThrAsnGlyIleTyrAsnThrLeuThrThrAsn...GluLeuSerG 1136
2870 GTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGAC 2919
1136 lySerGlyAsnPheTyrLeuHisThrAsnValAlaGlySerArgGlyAsp 1152
2920 AAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCAA 2969
1153 GlnLeuValValAsnAsnAsnAlaThrGlyAsnPheLysIlePheValGl 1169
2970 CAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAG 3019
1169 nAspThrGlyValSerProGlnSerAspAspAlaMetThrLeuVal.... 1184
3020 GGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAAC 3069
1185 LysThrGlyGlyGlyAspAlaSerPheThrLeuGlyAsn 1197
3070 GAA.....CACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAA 3113
1198 ThrGlyGlyPheValAspLeuGlyThrTyrGluTyr...ValLeuLysSe 1213
3114 AGACGGC.....GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGC 3157
1213 rAspGlyAsnSerAsnTrpAsnLeuThrAsnAspValLys..... 1226
3158 TTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGAC 3207
1226 1226
3208 AACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGA 3257
1227 Pr 1227
3258 AAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATG 3307
1227 oAsnProAspProIleProAsnPro..... 1235
3308 TCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAA 3357
1235 1235
3358 GACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCAC 3407
1236 LysProAspProLysProAspProLy 1244
3408 CGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGC 3457
1244 sProAspProAsnProLys.....ProAspProThrP 1255
3458 CCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTAT 3507
1255 roAspProThrProThrProValProGluLysArgIleThrPro..... 1269
3508 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTC.. 3555
1270 SerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheAs 1286
3556 GCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3592
1286 pAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSerP 1303
3593 ACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGNACACCAAACACTAC 3642
1303 roHisAsnAsnAsnValTrpGlyAlaThrTyrAsnThrArgAsnAsnVal 1319
3643 CGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAAT 3692
1320 ThrThrAspAlaGlyAlaGlyPheGluGlnThrLeuThrGlyMetThrVa 1336
3693 CGGTATGCAGAAAAAC.....CTCGGCAGCGGGCGCGTCGGCA 3730
1336 lGlyIleAspSerArgAsnAspIleProGluGlyIleThrThrLeuGlyA 1353
3731 TCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGC 3780
1353 laPheMetGlyTyrSerHisSerHisIleGlyPheAspArgGly..... 1367
3781 AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGG 3830
1368 GlyHisGlySerVal...GlySerTyrSerLeuGl 1378

```
3831 C..................................................... 3831
     |
1378 yGlyTyrAlaSerTrpGluHisGluSerGlyPheTyrLeuAspGlyValV 1395

3832 ..........AGGTTCGACATCGGCATC............AGCACGGGC 3858
               ||||||      ::::::           |||:::|||
1395 alLysLeuAsnArgPheLysSerAsnValAlaGlyLysMetSerSerGly 1411

3859 GCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCG 3908
     ::::::   :::::::::      |||:::|||:::||||||
1412 GlyAlaAlaAsnGlySerTyrHisSerAsnGlyLeuGlyGlyHis..... 1426

3909 CCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCG 3958
                          |||::::::      |||         |
1427 .....................IleGluThrGlyMetArgPheThrAspG 1436

3959 GCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAA 4008
     ||::::::::::::   ||||||   :::   |||   :::   ::::::
1436 lyAsnTrpAsnLeuThrProTyrAlaSerLeuThrGlyPheThrAlaAsp 1452

4009 AAAGCGGATTACCGCTACGAAAAC............GTCAATATCGCCAC 4046
     :::   :::|||:::   :::|||                 ::::::   ||
1453 AsnProGluTyrHisLeuSerAsnGlyMetLysSerLysSerValAspTh 1469

4047 CCCCGGTCTTGCGTTCAACCGNTACCGN...GCGGGCATTAAGGCAGATT 4093
     |   ::::::           ||||||      |||         :::|
1469 rArgSerIle............TyrArgGluLeuGlyAlaThrLeuSerT 1482

4094 ATTCATTCAAACCGGCGCAACACATNTCCATCACNCCTTATTTNAGCCTG 4143
     ||:::::::::   ::::::      ::::::   |||:::   :::
1482 yrAsnMetArgLeuGlyAsnGlyMetGluValGluProTrpLeuLysAla 1498

4144 TCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGC 4193
     :::                        |||::::::|||||||:::
1499 AlaValArgLysGluPheValAspAspAsnArgValLysValAsnSerAs 1515

4194 NGTATTGGCTCAGGATTTC......GGCAAAACCCGCAGTGCGGAATGGG 4237
                   ::::::      |||:::         :::   |
1515 pGlyAsnPheValAsnTyrLeuSerGlyArgArgGlyIleTyrGlnAlaG 1532

4238 GCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCNTCCACGCTGCCGCC 4287
     ||::::::|||:::::::::::::   |||||||||   |||   ::::::
1532 lyIleLysAlaSerPheSerSer...ThrLeuSerGlyHisLeuGlyVal 1547

4288 GCCAAAGGNCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAA..... 4332
     :::                       :::||||||||||||:::::
1548 Gly.................TyrSerHisSerAlaGlyValGluSerPr 1558

4333 .............TTAGGCTACCGCTGG 4347
                  |||   :::|||
1558 oTrpAsnAlaValAlaGlyValAsnTrp 1567
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505

seq_documentation_block:

```
ID   AAR63505 standard; Protein; 1536 AA.
XX
AC   AAR63505;
XX
DT   25-JUN-1995  (first entry)
XX
DE   Haemophilus high molecular weight protein HMW1.
XX
KW   High molecular weight protein; HMW1; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72293.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 2; Page 31; 127pp; English.
XX
CC   The HMW1 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
SQ   Sequence   1536 AA;
```

alignment_scores:

Quality: 282.00 Length: 1173
Ratio: 0.482 Gaps: 58
Percent Similarity: 49.872 Percent Identity: 20.460

alignment_block:
US-09-303-518D-651 x AAR63505 ..

Align seg 1/1 to: AAR63505 from: 1 to: 1536

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     |||:::|||||||:::|||                              :::::
 349 GlyGluThrTyrLeuGly.........................GlyAs 356

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
     :::::::|||:::          |||   |||   |||::::::   :::|
 356 pGluArgGlyGlu.........GlyLysLysGlyIleGlnLeuAlaLysL 370

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     ||                    |||:::   |||:::      |||:::
 370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
       |||      ::::::::::|||   |||   :::   ||||||
 381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

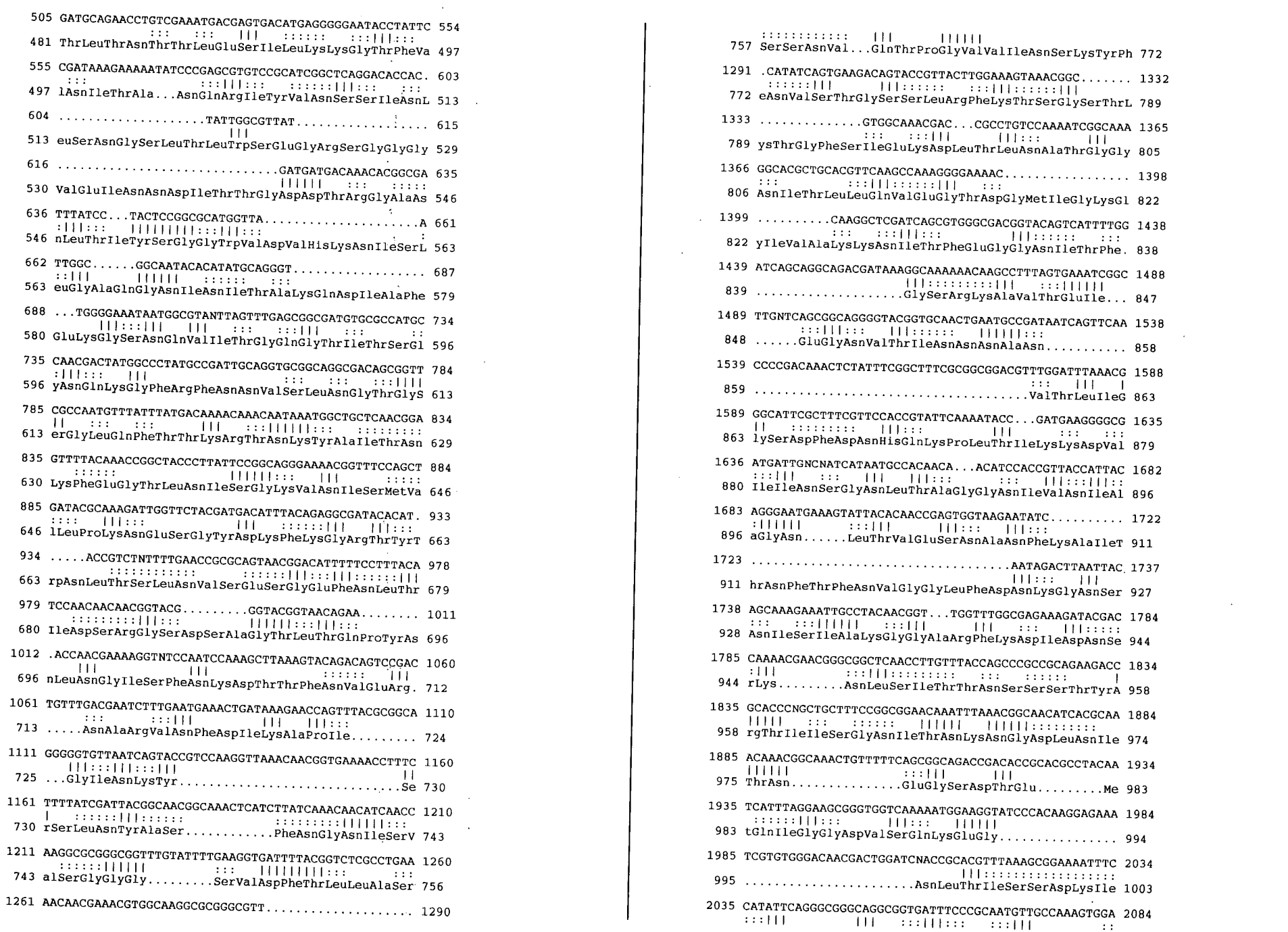
 327 TATTGTG..............................AGCGTGGCACATA 346
      ::::::                            :::::|||   :
 397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

 347 ACGGCGGCTAT..............................AACAAC 363
     ::|||||:::                              :::|||
 414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

 364 ......GTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTT 407
           ||||||                :::|||||:::      ::
 431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

 408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACC 457
     ::::  :::   :::   |||:::|||   :::   |||::::::::::
 447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

 458 CTTACAACGGCGAT...TANCATATGCCGCGTTTGCATAAATTTGTCACA 504
           :::|||:::   :::      |||:::   :::|||      |||
 464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480
```



```
1004 AsnIleThr...LysGlnIleThrIleLysAlaGlyValAspGlyGluAs 1019

2085 AGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2134
     ::::|||:::      :::|||:::|||:::
1019 nSerAspSerAspAlaThrAsnAsnAlaAsnLeu..............  1030

2135 CGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACA 2184
                  ||||||   |||:::
1031 ..............ThrIleLysThrLys.....................  1035

2185 AATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGAC 2234
           |||      :::|||:::|||      |||:::::::::::
1036 .........GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049

2235 TAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAA 2284
     :|||::::::   ::::::
1049 nLysAlaGluIleThrAlaLys............................  1056

2285 ANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTT...AGTGCAAATGGC 2331
          |||:::            ||||||      |||||||:::|||
1057 .....AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071

2332 GATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAG 2381
       :::         |||:::   |||   ::::::::::::::::||
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088

2382 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCA 2431
     |    |||                  :::::::::||||||:::::::
1088 rAlaAspGly.................HisLysValThrLeuHisSerL 1099

2432 ACNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCA 2481
     ::   ::::::|||||||:::   :::   |||   ::::::::::::
1099 ysValGluThrSerGlySerAsnAsn...AsnThrGluAspSerSerAsp 1114

2482 CAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCA 2531
     :::|||:::::::|||||||:::      :::||||||   |||||||:::
1115 AsnAsnAlaGlyLeuThrIle......AspAlaLys...AsnValThr.. 1127

2532 TTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2581
           :::|||:::|||::::::   :::   |||||||||
1128 .......ValAsnAsnAsnIleThr...SerHisLysAlaVal....... 1138

2582 TTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCA 2631
       :::   |||      :::|||:::::::::   :::      |||:::
1139 ..SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyThrThr 1154

2632 TTACACTTAAAA......GACAGCGAATGGACGCTGCCGTCAGGCACGGA 2675
     ::::::            :::   |||   |||      :::|||:::
1155 IleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySerIl 1171

2676 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2725
      ||||||::::::::   ::::::::::::::::||||||:::::::::
1171 eLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaThr. 1187

2726 ATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCG 2775
           ::::::|||   |||   ::::::      :::::::|||
1188 ........GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201

2776 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2825
           ::::::   :::   |||::::::      ::::::::
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIl 1218

2826 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAG 2875
     :::::::      ::::::   |||   :::   :::   :::      |
1218 eLysGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyG 1235

2876 GAACATTC.....................CGCTTTATGTCGGAACTCTTC 2904
     ||||||:::                    :::     :::::::|||
1235 lyThrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThr 1251

2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
           ::::::|||:::|||   |||   ::::::|||::::::
1252 ThrGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnVa 1267

2955 CACCTTGGCGGTC.....................AACAATACCGGCA 2980
       |||   |||:::                    :::|||||   |
1267 lThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValA 1284

2981 ACGAACCCGTAAGCCTCGATCAATTGACGGTA.................. 3012
     ||      ::::::     :::|||||||||
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIle 1300

3013 ......GTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTT 3056
           :::||||||     :::        |||:::::::
1301 AsnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuTh 1317

3057 CACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCA 3106
       |||      :::::::|||:::   ::: |||:::|||   :::|||
1317 rThrGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeu 1333

3107 TCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAG 3156
     |||||||:::   :::|||   |||      |||::::::   |||   ::
1334 SerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValTh 1350

3157 CTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGA 3206
     ::::   |||:::::::::            |||:::   :::   ::::
1350 rLeuAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnA 1367

3207 CAACGCGCAAAGCCTTG 3223
     ::|||      :::|||
1367 laThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846

seq_documentation_block:

```
ID   AAB01846 standard; Protein; 1536 AA.
XX
AC   AAB01846;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52195.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 28A-Q; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
```

CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 1536 AA;

alignment_scores:
Quality: 282.00 Length: 1173
Ratio: 0.482 Gaps: 58
Percent Similarity: 49.872 Percent Identity: 20.460

alignment_block:
US-09-303-518D-651 x AAB01846 ..

Align seq 1/1 to: AAB01846 from: 1 to: 1536

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     |||:::||||||:::|||                                 :::::
 349 GlyGluThrTyrLeuGly.........................GlyAs 356

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
     :::::::|||:::         |||   |||   |||::::::   :::|
 356 pGluArgGlyGlu.........GlyLysLysGlyIleGlnLeuAlaLysL 370

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     ||                    |||:::   |||:::      |||:::
 370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||      ::::::::::|||   |||   :::   ||||||
 381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

 327 TATTGTG............................AGCGTGGCACATA 346
      ::::::                          ::::::|||   :
 397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

 347 ACGGCGGCTAT..............................AACAAC 363
     ::|||||||:::                            :::|||
 414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

 364 ......GTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTT 407
           ||||||                 :::|||||||:::      ::
 431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

 408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACC 457
     ::::   :::   :::   |||:::|||   :::   |||::::::::::
 447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

 458 CTTACAACGGCGAT...TANCATATGCCGCGTTTGCATAAATTTGTCACA 504
         :::|||:::   :::      |||:::   :::|||      |||
 464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

 505 GATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATACCTATTC 554
                 :::   :::   |||   ::::::   :::|||:::
 481 ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheVa 497

 555 CGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC. 603
      :::                 :::|||:::   :::::::|||:::   :::
 497 lAsnIleThrAla...AsnGlnArgIleTyrValAsnSerSerIleAsnL 513

 604 ...................TATTGGCGTTAT..................... 615
                              |||
 513 euSerAsnGlySerLeuThrLeuTrpSerGluGlyArgSerGlyGlyGly 529

 616 ............................GATGATGACAAACACGGCGA 635
                                 ||||||   :::  :::::
 530 ValGluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAs 546

 636 TTTATCC...TACTCCGGCGCATGGTTA.....................A 661
     :|||:::   |||||||||:::|||:::                     :
 546 nLeuThrIleTyrSerGlyGlyTrpValAspValHisLysAsnIleSerL 563

 662 TTGGC......GGCAATACACATATGCAGGGT................... 687
     ::|||      ||||||   ::::::   :::
 563 euGlyAlaGlnGlyAsnIleAsnIleThrAlaLysGlnAspIleAlaPhe 579

 688 ...TGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGC 734
           |||:::|||   |||   :::   :::|||   :::      ::
 580 GluLysGlySerAsnGlnValIleThrGlyGlnGlyThrIleThrSerGl 596

 735 CAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTT 784
     :|||:::   |||                  :::   :::   :::||||
 596 yAsnGlnLysGlyPheArgPheAsnAsnValSerLeuAsnGlyThrGlyS 613

 785 CGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGA 834
     ||   :::   :::      |||   :::|||||:::   :::::::::
 613 erGlyLeuGlnPheThrThrLysArgThrAsnLysTyrAlaIleThrAsn 629

 835 GTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCT 884
      ::::::                  ||||||:::   |||      :::::
 630 LysPheGluGlyThrLeuAsnIleSerGlyLysValAsnIleSerMetVa 646

 885 GATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACAT. 933
     ::::   |||:::          |||   :::::|||   |||:::
 646 lLeuProLysAsnGluSerGlyTyrAspLysPheLysGlyArgThrTyrT 663

 934 .....ACCGTCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACA 978
           ::::::::::::      ::::::|||:::|||::::::|||
 663 rpAsnLeuThrSerLeuAsnValSerGluSerGlyGluPheAsnLeuThr 679

 979 TCCAACAACAACGGTACG.........GGTACGGTAACAGAA........ 1011
       :::::::::|||:::          ||||||:::|||:::
 680 IleAspSerArgGlySerAspSerAlaGlyThrLeuThrGlnProTyrAs 696

1012 .ACCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGAC 1060
        |||         |||                 ::::::   |||
 696 nLeuAsnGlyIleSerPheAsnLysAspThrThrPheAsnValGluArg. 712

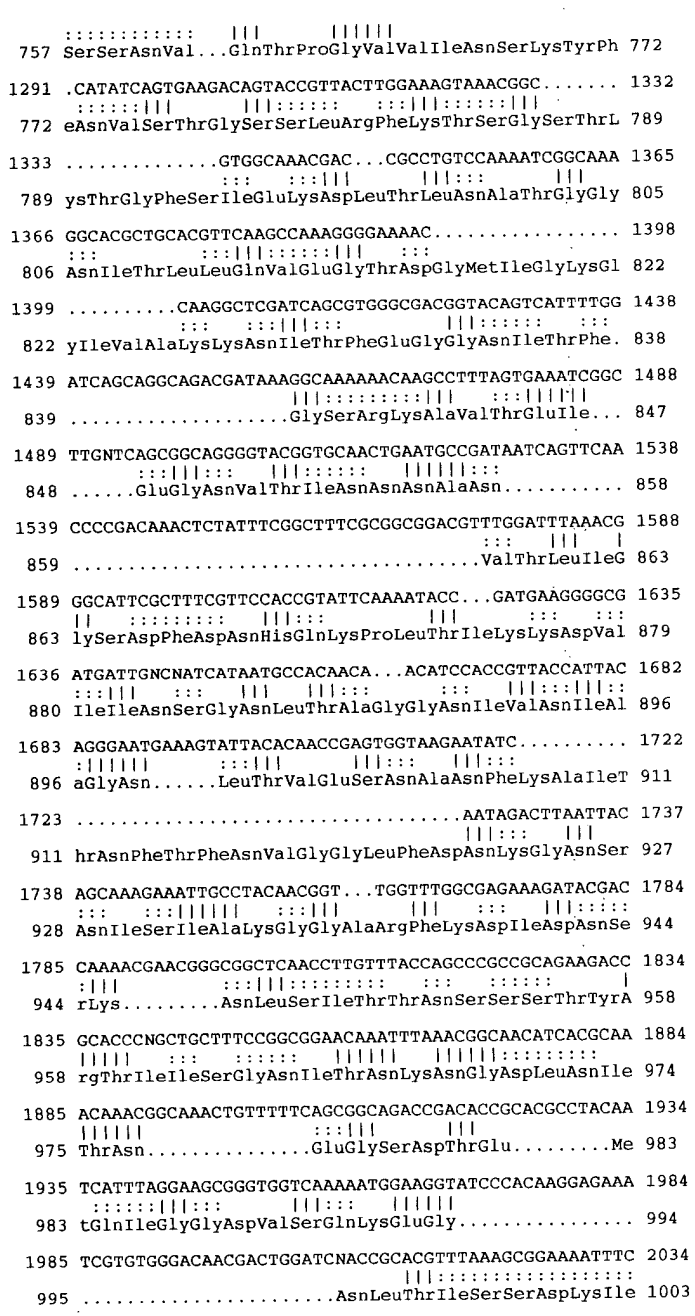
1061 TGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCGGCA 1110
        :::      :::|||         |||   |||:::
 713 .....AsnAlaArgValAsnPheAspIleLysAlaProIle......... 724

1111 GGGGGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTC 1160
       |||:::|||:::|||                              ||
 725 ...GlyIleAsnLysTyr............................Se 730

1161 TTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACATCAACC 1210
     |   ::::::|||:::::::          :::::::::|||||:::
 730 rSerLeuAsnTyrAlaSer............PheAsnGlyAsnIleServ 743

1211 AAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAA 1260
       ::::::||||||         :::   |||||||||:::      :::
 743 alSerGlyGlyGly.........SerValAspPheThrLeuLeuAlaSer 756

1261 AACAACGAAACGTGGCAAGGCGCGGGCGTT..................... 1290
```


SerSerAsnVal...GlnThrProGlyValValIleAsnSerLysTyrPh 772
.CATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGC..... 1332
eAsnValSerThrGlySerSerLeuArgPheLysThrSerGlySerThrL 789
.....GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAA 1365
ysThrGlyPheSerIleGluLysAspLeuThrLeuAsnAlaThrGlyGly 805
GGCACGCTGCACGTTCAAGCCAAAGGGGAAAAC..... 1398
AsnIleThrLeuLeuGlnValGluGlyThrAspGlyMetIleGlyLysGl 822
.....CAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGG 1438
yIleValAlaLysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe. 838
ATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGC 1488
.....GlySerArgLysAlaValThrGluIle... 847
TTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAA 1538
......GluGlyAsnValThrIleAsnAsnAsnAlaAsn..... 858
CCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACG 1588
.....ValThrLeuIleG 863
GGCATTCGCTTTCGTTCCACCGTATTCAAAATACC...GATGAAGGGGCG 1635
lySerAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspVal 879
ATGATTGNCNATCATAATGCCACAACA...ACATCCACCGTTACCATTAC 1682
IleIleAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAl 896
AGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATC..... 1722
aGlyAsn.....LeuThrValGluSerAsnAlaAsnPheLysAlaIleT 911
.....AATAGACTTAATTAC 1737
hrAsnPheThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSer 927
AGCAAAGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAGATACGAC 1784
AsnIleSerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSe 944
CAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACC 1834
rLys.....AsnLeuSerIleThrThrAsnSerSerSerThrTyrA 958
GCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAA 1884
rgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAsnIle 974
ACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAA 1934
ThrAsn.....GluGlySerAspThrGlu.....Me 983
TCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAA 1984
tGlnIleGlyGlyAspValSerGlnLysGluGly..... 994
TCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTC 2034
.....AsnLeuThrIleSerSerAspLysIle 1003
CATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGA 2084
1004 AsnIleThr...LysGlnIleThrIleLysAlaGlyValAspGlyGluAs 1019
2085 AGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2134
1019 nSerAspSerAspAlaThrAsnAsnAlaAsnLeu..... 1030
2135 CGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACA 2184
1031 ThrIleLysThrLys..... 1035
2185 AATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGAC 2234
1036 GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049
2235 TAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAA 2284
1049 nLysAlaGluIleThrAlaLys..... 1056
2285 ANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTT...AGTGCAAATGGC 2331
1057 AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071
2332 GATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAG 2381
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088
2382 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCA 2431
1088 rAlaAspGly.....HisLysValThrLeuHisSerL 1099
2432 ACNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCA 2481
1099 ysValGluThrSerGlySerAsnAsn...AsnThrGluAspSerSerAsp 1114
2482 CAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCA 2531
1115 AsnAsnAlaGlyLeuThrIle.....AspAlaLys...AsnValThr.. 1127
2532 TTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2581
1128 ValAsnAsnAsnIleThr...SerHisLysAlaVal..... 1138
2582 TTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCA 2631
1139 ..SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyThrThr 1154
2632 TTACACTTAAAA.....GACAGCGAATGGACGCTGCCGTCAGGCACGGA 2675
1155 IleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySerIl 1171
2676 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2725
1171 eLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaThr. 1187
2726 ATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCG 2775
1188 GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201
2776 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2825
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIl 1218
2826 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAG 2875
1218 eLysGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyG 1235
2876 GAACATTC.....CGCTTTATGTCGGAACTCTTC 2904
1235 lyThrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThr 1251
2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
1252 ThrGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnVa 1267

```
2955 CACCTTGGCGGTC.....................AACAATACCGGCA 2980
1267 lThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValA 1284

2981 ACGAACCCGTAAGCCTCGATCAATTGACGGTA..................  3012
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIle 1300

3013 ......GTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTT 3056
1301 AsnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuTh 1317

3057 CACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCA 3106
1317 rThrGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeu 1333

3107 TCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAG 3156
1334 SerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValTh 1350

3157 CTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGA 3206
1350 rLeuAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnA 1367

3207 CAACGCGCAAAGCCTTG 3223
1367 laThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293

seq_documentation_block:

```
ID   AAW30293 standard; Protein; 1536 AA.
XX
AC   AAW30293;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW1.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW1; hmw1A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by CTA"
FT   Misc-difference 98
FT                   /note= "encoded by GAT"
FT   Misc-difference 363
FT                   /note= "encoded by AAG"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90994 and AAT90996.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 66-70; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW1 (125 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW1 amino acid sequence was deduced from the
CC   hmw1 gene sequence (see AAT90994 and AAT90996).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW2 (see AAW30294), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW1
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW1 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1536 AA;
```

alignment_scores:

```
            Quality:  277.00            Length:   1173
              Ratio:  0.474               Gaps:     58
 Percent Similarity:  49.872   Percent Identity: 20.375
```

alignment_block:

US-09-303-518D-651 x AAW30293 ..

Align seq 1/1 to: AAW30293 from: 1 to: 1536

```
127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
349 GlyGluThrTyrLeuGly.........................GlyAs 356

177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

327 TATTGTG............................AGCGTGGCACATA 346
397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

347 ACGGCGGCTAT...............................AACAAC 363
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

364 ......GTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTT 407
431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

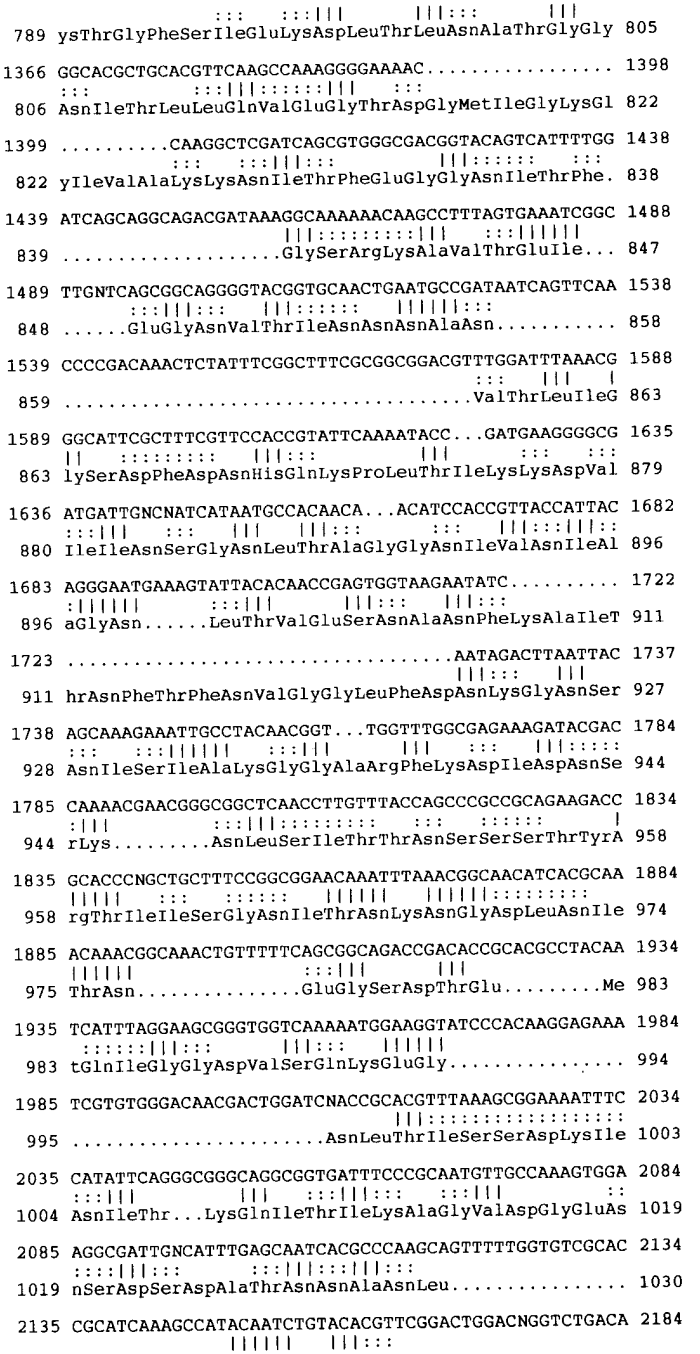
408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACC 457
447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

458 CTTACAACGGCGAT...TANCATATGCCGCGTTTGCATAAATTTGTCACA 504
464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

505 GATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATACCTATTC 554
481 ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheVa 497

555 CGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC. 603
497 lAsnIleThrAla...AsnGlnArgIleTyrValAsnSerSerIleAsnL 513
```





```
1031 .............ThrIleLysThrLys..................... 1035

2185 AATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGAC 2234
            |||      :::|||:::|||      |||:::::::::::
1036 .........GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049

2235 TAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAA 2284
     :|||:::::::   ::::::
1049 nLysAlaGluIleThrAlaLys........................... 1056

2285 ANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTT...AGTGCAAATGGC 2331
           |||:::            ||||||     |||||||:::|||
1057 .....AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071

2332 GATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAG 2381
     :::          |||:::   |||   ::::::::::::::::::||
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088

2382 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCA 2431
     |      |||                 :::::::::||||||:::::::
1088 rAlaAspGly.................HisLysValThrLeuHisSerL 1099

2432 ACNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCA 2481
     ::   ::::::||||||:::   :::   |||   ::::::::::::
1099 ysValGluThrSerGlySerAsnAsn...AsnThrGluHisSerSerAsp 1114

2482 CAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCA 2531
     :::|||::::::|||||||:::      :::|||||||   |||||||:::
1115 AsnAsnAlaGlyLeuThrIle......AspAlaLys...AsnValThr.. 1127

2532 TTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2581
          :::|||:::|||::::::   :::   |||||||||
1128 .......ValAsnAsnAsnIleThr...SerHisLysAlaVal....... 1138

2582 TTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCA 2631
     :::   |||      :::|||:::::::::   :::      |||:::
1139 ..SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyThrThr 1154

2632 TTACACTTAAAA......GACAGCGAATGGACGCTGCCGTCAGGCACGGA 2675
     ::::::            :::   |||   |||      :::|||:::
1155 IleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySerIl 1171

2676 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2725
      ||||||:::::::::   ::::::::::::::::|||||:::::::::
1171 eLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaThr. 1187

2726 ATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCG 2775
          :::::::|||   |||   ::::::      ::::::|||
1188 ........GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201

2776 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2825
          ::::::   :::   |||::::::         ::::::::
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIl 1218

2826 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAG 2875
     :::::::      ::::::   |||   :::   :::   :::      |
1218 eLysGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyG 1235

2876 GAACATTC.....................CGCTTTATGTCGGAACTCTTC 2904
     ||||||::                    :::      :::::::|||
1235 lyThrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThr 1251

2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
          ::::::|||:::|||   |||   ::::::|||::::::
1252 ThrGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnVa 1267

2955 CACCTTGGCGGTC.....................AACAATACCGGCA 2980
     |||   |||:::                    :::|||||   |
1267 lThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValA 1284

2981 ACGAACCCGTAAGCCTCGATCAATTGACGGTA.................. 3012
     ||      ::::::      :::|||||||||
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIle 1300

3013 ......GTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTT 3056
           :::|||||||      :::         |||::::::
1301 AsnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuTh 1317

3057 CACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCA 3106
      |||      :::::::|||:::   ::: |||:::|||   :::|||
1317 rThrGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeu 1333

3107 TCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAG 3156
     ||||||:::   :::|||   |||      |||::::::   |||   ::
1334 SerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValTh 1350

3157 CTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGA 3206
     ::::   |||::::::::::            |||:::   :::   ::::
1350 rLeuAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnA 1367

3207 CAACGCGCAAAGCCTTG 3223
     ::|||      :::|||
1367 laThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723

seq_documentation_block:

ID AAR41723 standard; Protein; 1536 AA.
XX
AC AAR41723;
XX
DT 26-APR-1994 (first entry)
XX
DE High molecular weight protein 1 (HMW1).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus influenzae.
XX
PN WO9319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Barenkamp SJ;
XX
DR WPI; 1993-320683/40.
DR N-PSDB; AAQ49506.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX
PS Claim 3; Figure 2; 100pp; English.
XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
SQ Sequence 1536 AA;

alignment_scores:

Quality: 271.00 Length: 1185
Ratio: 0.465 Gaps: 58
Percent Similarity: 49.198 Percent Identity: 20.084

alignment_block:

US-09-303-518D-651 x AAR41723 ..

Align seg 1/1 to: AAR41723 from: 1 to: 1536

```
127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
    |||:::|||||||:::|||                              :::::
349 GlyGluThrTyrLeuGly.....................................GlyAs 356

177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
    :::::::|||:::          |||   :::   |||:::::::   :::|
356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
    ||                    |||:::   |||:::      |||:::
370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
      |||      ::::::::::|||   |||   :::   ||||||
381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

327 TATTGTG................................AGCGTGGCACATA 346
     ::::::                              ::::::|||   :
397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

347 ACGGCGGC.......................................... 354
    ::|||||||
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

355 ..............................TATAACAACGTTGATTT 371
                                  ::::::||||||:::::
431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

372 TGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATTG 421
    ::::||||||   ::::::
447 eAsnAlaGluThrAlaGly.............................. 453

422 TGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGGCGAT 471
        |||:::|||   :::   |||:::::::::      :::|||:::
454 .....ArgSerAsnThrSerGluAspAspGluTyrThrGlySerGlyAsn 468

472 ...TANCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGT 518
     :::      |||:::   :::|||      |||           ::
469 SerAlaSerThrProLysArgAsnLysGluLysThrThrLeuThrAsnTh 485

519 CGAAATGACGAGTGACATGAGGGGGAATACCTATTCCGATAAAGAAAAAT 568
    :  :::   |||   ::::::   :::|||:::   :::
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla. 501

569 ATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC.............. 603
       :::|||:::   :::::|||:::   :::
502 ..AsnGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySer 517

604 ......TATTGGCGTTAT............................... 615
           |||
518 LeuThrLeuTrpSerGluGlyArgSerGlyGlyGlyValGluIleAsnAs 534

616 ..............GATGATGACAAACACGGCGATTTATCC...TACT 646
                |||||   :::   ::::::|||:::   ||||
534 nAspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyrS 551

647 CCGGCGCATGGTTA.....................ATTGGC......GGC 669
    |||||:::|||:::                     :::|||      |||
551 erGlyGlyTrpValAspValHisLysAsnIleSerLeuGlyAlaGlnGly 567

670 AATACACATATGCAGGGT.....................TGGGGAAATAA 698
    |||   ::::::   :::                     |||:::||
568 AsnIleAsnIleThrAlaLysGlnAspIleAlaPheGluLysGlySerAs 584

699 TGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCC 748
    |   |||   :::   :::|||   :::      :::|||:::   |||
584 nGlnValIleThrGlyGlnGlyThrIleThrSerGlyAsnGlnLysGlyP 601

749 CTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATT 798
                                  :::   :::   :::|||||   :::   :::
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

799 TATGACAAAACAAACAATAAATGGCTGCTCAACGGAGTTTTACAAACCGG 848
        |||   :::|||||:::   :::::::::   ::::::
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

849 CTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGCAAAGATT 898
              ||||||:::   |||      :::::::::   |||:::
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

899 GGTTCTACGATGACATTTACAGAGGCGATACACAT......ACCGTCTNT 942
          |||   ::::::|||   |||:::          ::::::
651 luSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSer 667

943 TTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGG 992
    ::::::      ::::::|||:::|||:::::::|||   :::::::::||
668 LeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgGl 684

993 TACG.........GGTACGGTAACAGAA.........ACCAACGAAAAGG 1024
    |:::         ||||||:::|||:::             |||
684 ySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIleS 701

1025 TNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCT 1074
         |||                ::::::   |||      :::
701 erPheAsnLysAspThrThrPheAsnValGluArg......AsnAlaArg 715

1075 TTGAATGAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCA 1124
     :::|||         |||   |||:::             |||:::|||::
716 ValAsnPheAspIleLysAlaProIle............GlyIleAsnLy 728

1125 GTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTACG 1174
     :|||                                  |||   ::::::|||:
728 sTyr.............................SerSerLeuAsnTyrA 735

1175 GCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGT 1224
     :::::          :::::::::|||||||:::   ::::::||||||
735 laSer.............PheAsnGlyAsnIleSerValSerGlyGlyGly 747

1225 TTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTG 1274
               :::   ||||||||||:::      :::::::::::::::
748 .........SerValAspPheThrLeuLeuAlaSerSerSerAsnVal.. 760

1275 GCAAGGCGCGGGCGTT.....................CATATCAGTGAAG 1303
      |||      ||||||                    :::::::|||
761 .GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerThrG 777

1304 ACAGTACCGTTACTTGGAAAGTAAACGGC..................... 1332
       |||::::::   :::|||:::::::|||
777 lySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPheSer 793

1333 GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGT 1379
     :::   :::|||      |||:::      |||   :::         ::
794 IleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLeuLe 810

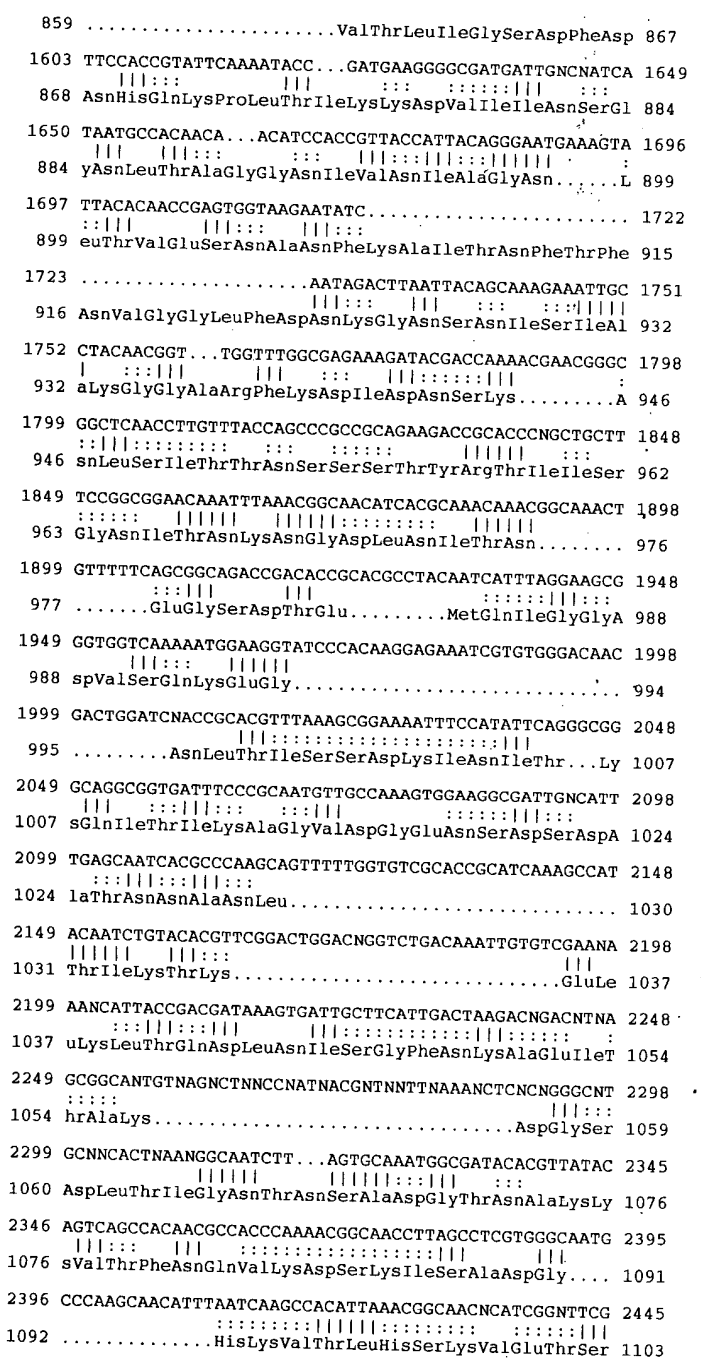
1380 TCAAGCCAAAGGGGAAAAC.........................CAAG 1402
     :|||:::::::|||   :::                        :::
810 uGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaLysL 827

1403 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1452
       :::|||:::         |||::::::   :::
827 ysAsnIleThrPheGluGlyGlyAsnIleThrPhe.............. 838

1453 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAG 1502
           |||:::::::::|||   :::|||||         :::|||::
839 ......GlySerArgLysAlaValThrGluIle.........GluGlyAs 850

1503 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1552
     :   |||:::::::   ||||||:::                        858
850 nValThrIleAsnAsnAsnAlaAsn.......................... 858

1553 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1602
                                  :::   |||   |||   :::::::::
```


859 ValThrLeuIleGlySerAspPheAsp 867
1603 TTCCACCGTATTCAAAATACC...GATGAAGGGGCGATGATTGNCNATCA 1649
868 AsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGl 884
1650 TAATGCCACAACA...ACATCCACCGTTACCATTACAGGGAATGAAAGTA 1696
884 yAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsn.....L 899
1697 TTACACAACCGAGTGGTAAGAATATC..... 1722
899 euThrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPhe 915
1723 AATAGACTTAATTACAGCAAAGAAATTGC 1751
916 AsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAl 932
1752 CTACAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGC 1798
932 aLysGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....A 946
1799 GGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTT 1848
946 snLeuSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSer 962
1849 TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACT 1898
963 GlyAsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1899 GTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCG 1948
977 GluGlySerAspThrGlu.....MetGlnIleGlyGlyA 988
1949 GGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAAC 1998
988 spValSerGlnLysGluGly..... 994
1999 GACTGGATCNACCGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGG 2048
995 AsnLeuThrIleSerSerAspLysIleAsnIleThr...Ly 1007
2049 GCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGNCATT 2098
1007 sGlnIleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspA 1024
2099 TGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCAT 2148
1024 laThrAsnAsnAlaAsnLeu..... 1030
2149 ACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANA 2198
1031 ThrIleLysThrLys.....GluLe 1037
2199 AANCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACNGACNTNA 2248
1037 uLysLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleT 1054
2249 GCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNT 2298
1054 hrAlaLys.....AspGlySer 1059
2299 GCNNCACTNAANGGCAATCTT...AGTGCAAATGGCGATACACGTTATAC 2345
1060 AspLeuThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLy 1076
2346 AGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATG 2395
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspGly.... 1091
2396 CCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACNCATCGGNTTCG 2445
1092 HisLysValThrLeuHisSerLysValGluThrSer 1103
2446 GGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGTCT 2495
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLe 1119
2496 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2545
1119 uThrIle.....AspAlaLys...AsnValThr.....ValAsnA 1130
2546 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGC 2595
1130 snAsnIleThr...SerHisLysAlaVal.....SerIleSerAla 1142
2596 TTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTAAAA.. 2643
1143 ThrSerGlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrTh 1159
2644 GACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAA 2689
1159 rGlyAsnValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleG 1176
2690 ACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCT 2739
1176 luSerSerSerGlySerValThrLeuThrAlaThr.....GluGly 1189
2740 GCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCG 2789
1190 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAs 1206
2790 CCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCA 2839
1206 nSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGlyThrGluS 1223
2840 ACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTC..... 2883
1223 erValThrThrSerSerGlnSerGlyAspIleGlyGlyThrIleSerGly 1239
2884 CGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGA 2918
1240 GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAsnSe 1256
2919 CAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTC. 2967
1256 rLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaThrG 1272
2968 AACAATACCGGCAACGAACCCGTAAGC 2994
1272 lyThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsn 1288
2995 CTCGATCAATTGACGGTA.....GTGGAAGG 3020
1289 AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGl 1305
3021 GAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACG 3070
1305 yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerS 1322
3071 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3120
1322 erHisIleThrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGl 1338
3121 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3170
1338 ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrG 1355
3171 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCC 3220
1355 lyThrLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371
3221 TTG 3223
1372 Leu 1372

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725

seq_documentation_block:
ID   AAR41725 standard; Protein; 1536 AA.
XX
AC   AAR41725;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 1 (HMW1).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 668..677
FT                   /note= "Possibly incorrect sequence. Alternative
FT                          sequence for this region is LNVSESGEFN.
FT                          (See comments)."
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49508.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 3; Figure 2/10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41723 (High molecular weight protein 1)
CC   although it does differ slightly. (Repeated regions which are
CC   possibly incorrect and occur in the corresponding nucleotide coding
CC   sequence contribute to these differences).
XX
SQ   Sequence   1536 AA;


alignment_scores:
            Quality:  265.00              Length:   1173
              Ratio:  0.457                 Gaps:     55
 Percent Similarity:  49.446   Percent Identity: 20.119

alignment_block:
 US-09-303-518D-651 x AAR41725   ..

 Align seg 1/1  to: AAR41725  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::||||||:::|||                              :::::
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
         :::::::|||:::         |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                        |||:::   |||:::      |||:::
     370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||      ::::::::::|||   |||   :::   ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG............................AGCGTGGCACATA 346
         ::::::                            ::::::|||   :
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGCTAT..............................AACAAC 363
         ::||||||:::                                ::|||
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     364 ......GTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTT 407
               ||||||                  :::|||||:::      ::
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

     408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACC 457
         ::::   :::   :::   |||:::|||   :::   |||::::::::::
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     458 CTTACAACGGCGAT...TANCATATGCCGCGTTTGCATAAATTTGTCACA 504
             :::|||:::   :::      |||:::   :::|||      |||
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     505 GATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATACCTATTC 554
                     :::   :::   |||   ::::::   :::|||:::
     481 ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheVa 497

     555 CGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC. 603
         :::                 :::|||:::   :::::::|||:::   :::
     497 lAsnIleThrAla...AsnGlnArgIleTyrValAsnSerSerIleAsnL 513

     604 ...................TATTGGCGTTAT..................... 615
                             |||
     513 euSerAsnGlySerLeuThrLeuTrpSerGluGlyArgSerGlyGlyGly 529

     616 ..............................GATGATGACAAACACGGCGA 635
                                      ||||||   :::   :::::
     530 ValGluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAs 546

     636 TTTATCC...TACTCCGGCGCATGGTTA.....................A 661
         :|||:::   ||||||||||:::|||:::                   :
     546 nLeuThrIleTyrSerGlyGlyTrpValAspValHisLysAsnIleSerL 563

     662 TTGGC......GGCAATACACATATGCAGGGT................... 687
         ::|||      ||||||   ::::::   :::
     563 euGlyAlaGlnGlyAsnIleAsnIleThrAlaLysGlnAspIleAlaPhe 579

     688 ...TGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGC 734
             |||:::|||   |||   :::   ::::|||   :::       ::
     580 GluLysGlySerAsnGlnValIleThrGlyGlnGlyThrIleThrSerGl 596

     735 CAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTT 784
         :|||:::   |||                        :::   :::   :::|||||
     596 yAsnGlnLysGlyPheArgPheAsnAsnValSerLeuAsnGlyThrGlyS 613

     785 CGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGA 834
         ||   :::   :::       |||   :::||||||:::   :::::::::
     613 erGlyLeuGlnPheThrThrLysArgThrAsnLysTyrAlaIleThrAsn 629

     835 GTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCT 884
             ::::::                |||||||:::   |||       :::::
     630 LysPheGluGlyThrLeuAsnIleSerGlyLysValAsnIleSerMetVa 646

     885 GATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACAT. 933
         ::::   |||:::            |||   :::::::|||   |||:::
     646 lLeuProLysAsnGluSerGlyTyrAspLysPheLysGlyArgThrTyrT 663
```

```
 934 .....ACCGTCTNT 942
 663 rpAsnLeuThrSerLysValAspMetIleAsnSerLysAspAlaLeuThr 679

 943 TTTGAACCGCGC.....AGTAACGGACATTTTTCCTTTACATCCAA 983
 680 IleAspSerArgGlySerAspSerAlaGlyThrLeuThrGlnProTyrAs 696

 984 CAACAACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATC 1033
 696 nLeuAsnGlyIleSerPheAsnLysAspThrThrPheAsnValGluArg. 712

1034 CAAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAA 1083
 713 .....AsnAlaArgValAsnPhe 718

1084 ACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCC 1133
 719 AspIleLysAlaProIle.....GlyIleAsnLysTyr..... 729

1134 AAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCA 1183
 730 .....SerSerLeuAsnTyrAlaSer.... 736

1184 AACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1233
 737 .....PheAsnGlyAsnIleServalSerGlyGlyGly..... 747

1234 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGC 1283
 748 SerValAspPheThrLeuLeuAlaSerSerSerAsnVal...GlnThrPr 763

1284 GGGCGTT.....CATATCAGTGAAGACAGTACCG 1312
 763 oGlyValValIleAsnSerLysTyrPheAsnValSerThrGlySerSerL 780

1313 TTACTTGGAAAGTAAACGGC.....GTGGCAAAC 1341
 780 euArgPheLysThrSerGlySerThrLysThrGlyPheSerIleGluLys 796

1342 GAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAA 1388
 797 AspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGl 813

1389 AGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGG 1438
 813 uGlyThrAspGly.....MetIleGlyLysGlyIleValAlaLysL 827

1439 ATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGC 1488
 827 ysAsnIleThrPheGluGlyGlyLysMetArgPheGlySerArgLysAla 843

1489 TTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAA 1538
 844 ValThrGluIleGluGlyAsnValThrIleAsn..... 854

1539 CCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACG 1588
 855 .....AsnAsnAlaAsnValThrLeuIleG 863

1589 GGCATTCGCTTTCGTTCCACCGTATTCAAAATACC...GATGAAGGGGCG 1635
 863 lySerAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspVal 879

1636 ATGATTGNCNATCATAATGCCACAACA...ACATCCACCGTTACCATTAC 1682
 880 IleIleAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAl 896

1683 AGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATC..... 1722
 896 aGlyAsn.....LeuThrValGluSerAsnAlaAsnPheLysAlaIleT 911

1723 .....AATAGACTTAATTAC 1737
```

```
 911 hrAsnPheThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSer 927

1738 AGCAAAGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAGATACGAC 1784
 928 AsnIleSerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSe 944

1785 CAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACC 1834
 944 rLys.....AsnLeuSerIleThrThrAsnSerSerSerThrTyrA 958

1835 GCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAA 1884
 958 rgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAsnIle 974

1885 ACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAA 1934
 975 ThrAsn.....GluGlySerAspThrGlu.....Me 983

1935 TCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAA 1984
 983 tGlnIleGlyGlyAspValSerGlnLysGluGly..... 994

1985 TCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTC 2034
 995 .....AsnLeuThrIleSerSerAspLysIle 1003

2035 CATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGA 2084
1004 AsnIleThr...LysGlnIleThrIleLysAlaGlyValAspGlyGluAs 1019

2085 AGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2134
1019 nSerAspSerAspAlaThrAsnAsnAlaAsnLeu..... 1030

2135 CGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACA 2184
1031 .....ThrIleLysThrLys..... 1035

2185 AATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGAC 2234
1036 .....GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049

2235 TAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAA 2284
1049 nLysAlaGluIleThrAlaLys..... 1056

2285 ANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTT...AGTGCAAATGGC 2331
1057 .....AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071

2332 GATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAG 2381
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088

2382 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCA 2431
1088 rAlaAspGly.....HisLysValThrLeuHisSerL 1099

2432 ACNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCA 2481
1099 ysValGluThrSerGlySerAsnAsn...AsnThrGluAspSerSerAsp 1114

2482 CAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCA 2531
1115 AsnAsnAlaGlyLeuThrIle.....AspAlaLys...AsnValThr.. 1127

2532 TTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2581
1128 .....ValAsnAsnAsnIleThr...SerHisLysAlaVal..... 1138

2582 TTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCA 2631
```

```
1139 ..SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyThrThr 1154

2632 TTACACTTAAAA......GACAGCGAATGGACGCTGCCGTCAGGCACGGA 2675
     ::::::            :::   |||   |||      :::|||:::
1155 IleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySerIl 1171

2676 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2725
     ||||||:::::::::   ::::::::::::::::||||||:::::::::
1171 eLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaThr. 1187

2726 ATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCG 2775
           ::::::|||   |||   ::::::      :::::::|||
1188 ........GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201

2776 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2825
           ::::::   :::   |||::::::          ::::::::
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIl 1218

2826 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAG 2875
     :::::::      ::::::   |||   :::   :::   :::     |
1218 eLysGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyG 1235

2876 GAACATTC.....................CGCTTTATGTCGGAACTCTTC 2904
     ||||||:::                   :::      ::::::|||
1235 lyThrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThr 1251

2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
           ::::::|||:::|||   |||   ::::::|||::::::
1252 ThrGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnVa 1267

2955 CACCTTGGCGGTC.......................AACAATACCGGCA 2980
      |||   |||:::                       :::||||||   |
1267 lThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValA 1284

2981 ACGAACCCGTAAGCCTCGATCAATTGACGGTA..................... 3012
     ||      ::::::     :::||||||||||
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIle 1300

3013 ......GTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTT 3056
           :::||||||      :::       |||::::::
1301 AsnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuTh 1317

3057 CACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCA 3106
      |||      :::::::|||:::   ::: |||:::|||   :::|||
1317 rThrGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeu 1333

3107 TCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAG 3156
     |||||||:::   :::|||   |||      |||:::::::   |||   ::
1334 SerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValTh 1350

3157 CTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGA 3206
     ::::   |||:::::::::            |||:::   :::   ::::
1350 rLeuAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnA 1367

3207 CAACGCGCAAAGCCTTG 3223
     ::|||      :::|||
1367 laThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824

seq_documentation_block:

ID AAB01824 standard; Protein; 1227 AA.
XX
AC AAB01824;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain Joyc HMW1A protein, SEQ ID NO:26.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain Joyc.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52175.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 18A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 1227 AA;

alignment_scores:

Quality: 264.50 Length: 1161
Ratio: 0.474 Gaps: 62
Percent Similarity: 48.062 Percent Identity: 21.102

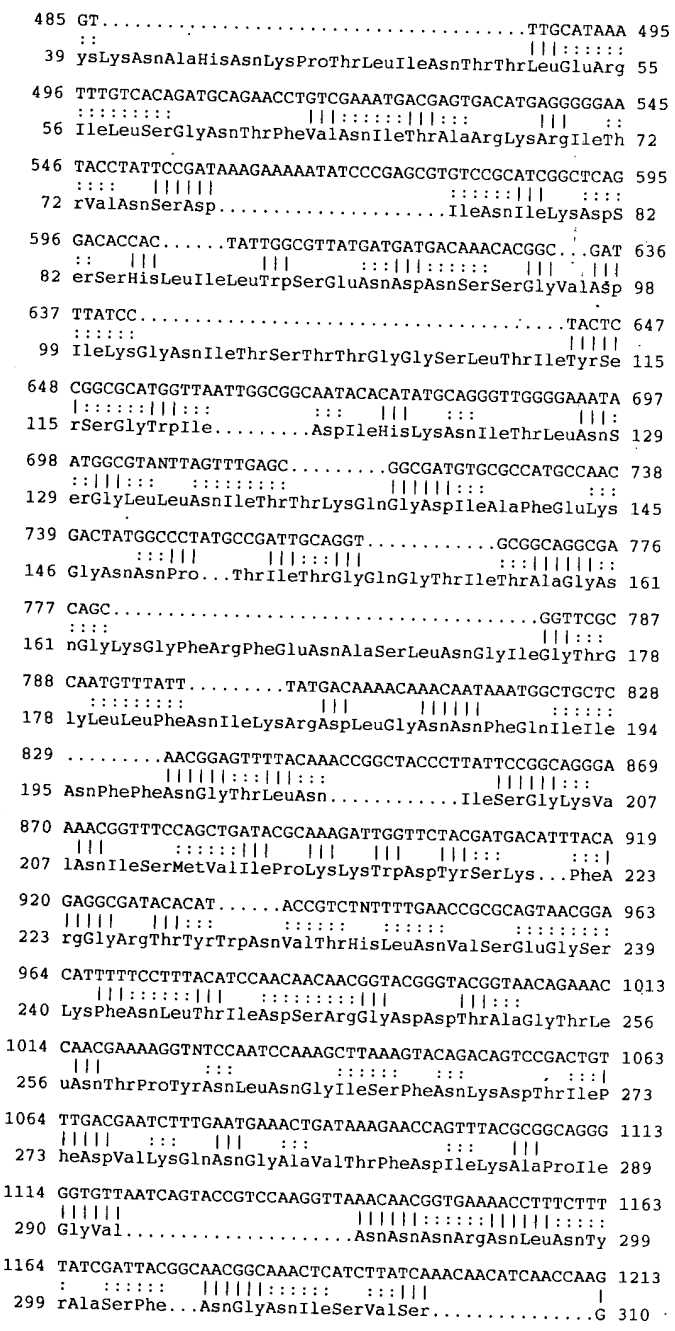
alignment_block:

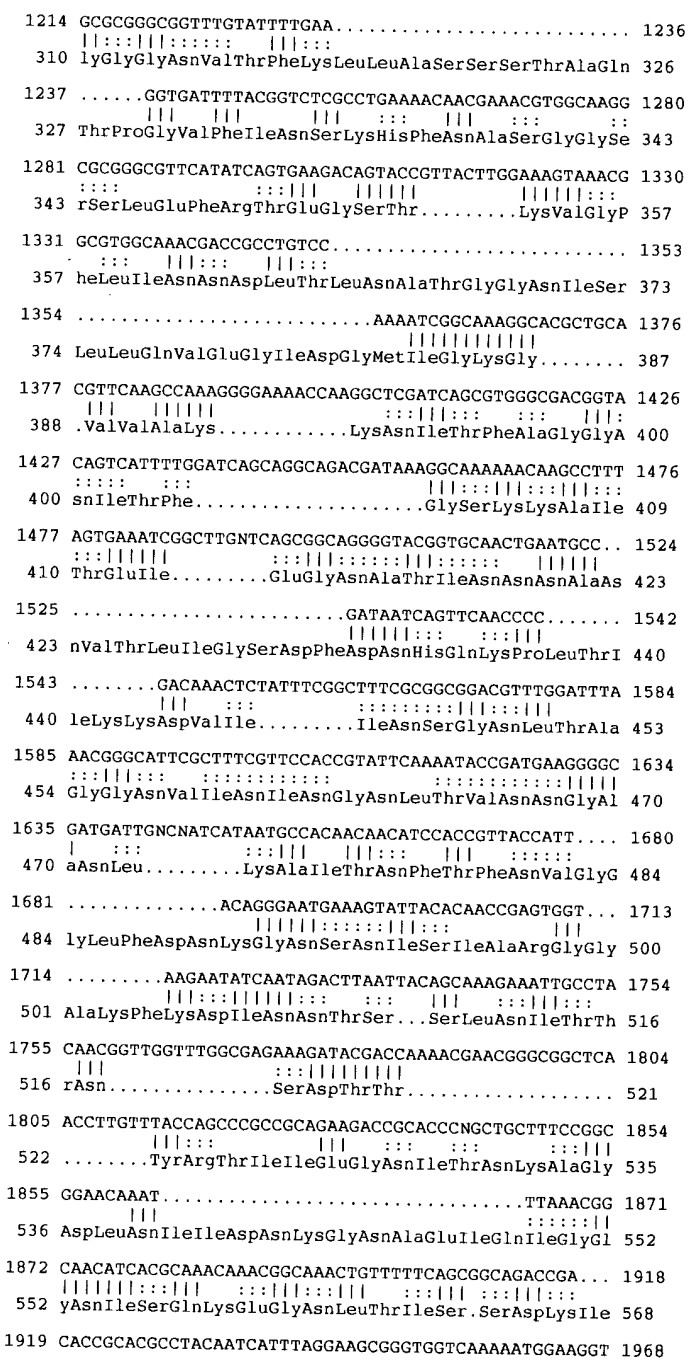
US-09-303-518D-651 x AAB01824 ..

Align seq 1/1 to: AAB01824 from: 1 to: 1227

```
 388 AATCCCGATCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTA 437
     :::|||||||:::      ::::::          :::|||||||:::
   6 AspProAspAsnValSerIleAsnAlaGlyThrSerGluArgAsnAspAl 22

 438 TAAGCCTGACAATTCACACCCTTACAAC...GGCGATTANCATATGCCGC 484
      :::|||   ::::::   |||   :::   |||           |||:
  22 aSerProThrGluAspPheProThrGlyAlaGlyGlyLysAspAsnProL 39
```





```
 569 AsnIleThrAsnGlnIleThrIleLysLysGly..... 579

1969 ATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGC....A 2014
     ::::::::::   :::   :::::::::::::::::::::      |
 580 .ValAsnLysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLeuT 596

2015 CGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCC 2064
     ||:::|||:::::::::::::::::::   :::          :::   ::: :
 596 hrIleLysThrLysGluLeuGlnLeuThrGlyAspLeuAsnIleSerGly 612

2065 CGCAATGTTGCCAAAGTGGAA.........................GG 2087
       :::   :::::::::                              ||
 613 PheAspLysAlaGluIleThrAlaLysGluGlyAlaAspLeuIleIleGl 629

2088 CGATTGNCATTTGAGCAATCACGCCCAAGCA.........GTTTTTGGTG 2128
     |:::::::      :::|||:::||||:::|||          :::||||::
 629 yAsnSerAspAsnAsnAsnAsnAlaAsnAlaLysLysValThrPheAsnG 646

2129 TCGCACCGCATCAA..............AGCCATACAATCTGTACACGT 2163
       :::      :::                ||||||:::::      :::
 646 lnValLysAspSerLysIleSerAlaGlySerHisAsnValThrLeuAsn 662

2164 TCG.............................................. 2166
     |||
 663 SerLysValGluThrSerAsnGlyAsnAsnAspAlaGluSerAsnAsnGl 679

2167 .GACTGGACNGGTCTGACA..............AATTGTGTCGAANAAA 2200
      |||   |||:::||||||                  |||:::
 679 yAspSerThrSerLeuThrIleAsnAlaLysAsnValThrValAsnAsnA 696

2201 NCATTACCGACGATAAAGTGATT.......................... 2223
       |||||||:::   |||::::::
 696 snIleThrSerHisLysThrValAsnIleThrAlaSerGluAsnValThr 712

2224 ......GCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNC 2267
           |||:::   |||   :::   :::   |||:::|||   :::::
 713 ThrLysAlaGlyThrThrIleAsnAlaThrIleGlySerValGluValTh 729

2268 CNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC.... 2313
     ::::   :::                |||      :::   ::::||
 729 rAlaLysThrGlyAspIleLysGlyGlyIleGluSerAsnSerGlyAsnV 746

2314 ..AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCC 2361
       |||::::::|||:::|||||||||      :::|||||||:: : ::
 746 alAsnIleThrAlaSerGlyAspThr...LeuAsnValSerAsnIleThr 761

2362 ACCCAAAACGGCAACCTT............................... 2379
       ||||||   ::::::
 762 GlyGlnAsnValThrValAlaAlaAlaSerGlyAlaValThrThrThrLy 778

2380 ..............AGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2413
                   :::   :::||||||||||:::   |||   :::
 778 sGlySerThrIleAsnAlaThrThrGlyAsnAlaAsnIleThrThrLysT 795

2414 AAGCCACATTAAACGGC.........................AACNCA 2436
       :::   :::||||||                          |||
 795 hrGlyGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIle 811

2437 TCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAA 2486
     ::::::||||||||||:::      |||:::||||||   :::::::|||
 812 ThrAlaSerGlyAsnThrLeuAsnAsnValSerAsnIleThrGlyGlnAs 828

2487 CGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCG 2536
     |   ::::::|||   ::::::::::|||   ::::::::::::   |||:
 828 nValThrValThrAlaAsnSerGlyAlaIleThrThrThrGluGlySerT 845

2537 CACTCAAC.........GGCAATGTCTCCCTAGCCGATAAGGCAGTATTC 2577
     :::::|||         |||:::::::::::::::   :::
 845 hrIleAsnAlaThrThrGlyAspAlaAsnIleThrThrGln......... 858

2578 CATTTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANAC 2627
                          ||||||::::::::::|||:::      ::
```

```
 859 .....................ThrGlyAsnIleAsnGlyLysValGluSe 868

2628 AGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAAT 2677
     ::::         :::   |||          :::   ::::|||      |
 868 rSer.........SerGlySerValThrLeuIleAlaThrGlyGlnThrL 882

2678 TA......GGCAATTTAAACCTTGACAACGCCACCATTACACTCAAT... 2718
     ||      ||||||::::::   |||::::::|||||||||   :::
 882 euAlaValGlyAsnIleSerGlyAspThrValThrIleThrAlaAspLys 898

2719 ................................TCCGCCTATCGCCACGA 2735
                                     ::::::   :::
 899 GlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThrLysSerVa 915

2736 TGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTT 2785
       ::::::::::::|||:::|||   :::|||   |||
 915 lThrThrSerSerGlnSerGlyAspIleSerGlyThr.............. 927

2786 CGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGT 2835
                                          ::::::::::::
 928 ....................................IleSerGlyAsn 931

2836 TTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCG 2885
           ::::::::::::::::|||:::|||::::::|||:::
 932 ThrValSerValSerAlaThrGlySerLeuThrThrGlnAla........ 945

2886 CTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGG 2935
                                     ::::::|||::::::   |||:
 946 ............................GlySerLysIleGlu...AlaL 952

2936 AAAGTTCCGAAGGNACTTACACCTTGGCGGTC.................. 2967
       :::::::::|||::::::   |||   |||:::
 952 ysThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGlyThr 968

2968 ......AACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGT 3011
             :::|||||||   |||      ::::::   |||:::|||||||::
 969 IleSerGlyAsnThrValAsnValThrAlaAsnThrAspAsnLeuThrIl 985

3012 AGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTC.... 3057
     :   :::|||          |||      :::      :::::::::
 985 eLysAspGlyAlaArgIleLysAlaThrGlyGlyAlaValThrLeuThrA 1002

3058 ...........ACCCTGCAAAACGAACACGTCG.........ATGCCGGC 3087
                 ||||||    |||:::||||||                ::
1002 laThrGlyGlyThrLeu.ThrThrGluThrSerSerAspIleThrSerSe 1018

3088 GCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAA 3137
     ::::|||   |||:::   :::||||||   :::||||   |||      |
1018 rAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIleAlaGlySerI 1035

3138 TCCGGTCAAAGAACAAGAGCTTTCCGACA 3166
     ||:::::::   |||   ::::::   |||
1035 leAsnAlaAlaAsnValThrLeuAsnThr 1044
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830

seq_documentation_block:
ID AAB01830 standard; Protein; 1222 AA.
XX
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)
XX
DE H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain K1.
XX

```
FH   Key             Location/Qualifiers
FT   Misc-difference 307
FT                   /note= "Encoded by GG"
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52180.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 20A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1222 AA;
```

```
alignment_scores:
            Quality:  264.00              Length:   1325
              Ratio:  0.401                Gaps:     60
 Percent Similarity:  49.736   Percent Identity: 20.377

alignment_block:
 US-09-303-518D-651 x AAB01830   ..

 Align seg 1/1  to: AAB01830  from: 1  to: 1222

     346 AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGA 395
         |||   :::   |||          :::|||         ::::::|||::
       3 AsnValSerIleAsnAlaProAlaLeuGlyArgThrGluSerThrProAs 19

     396 TCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTG 445
         :::::::      :::|||:::      ::::::   |||||||||
      19 nAsnAsn......GluTyrAspSerProAsnGlnIleAsnTyrLysAsnL 34

     446 ACAATTCACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAA 495
           |||                    ::::::   :::   |||:::   |||
      34 ysProSerLeuSerThrLeuThrAsnThrThrLeuGluArgIleLeuLys 50

     496 TTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAA 545
           |||:::         |||:::::::|||:::             ::
      51 ArgAsnThrSer.........ValAsnIleThrAlaThrLysThrIleTh 64

     546 TACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAG 595
         ::::   ||||||                        ::::::||||||::::
      64 rValAsnSerAsp.....................IleAsnIleGlyAspS 74

     596 GACACCACTATTGGCGTTATGATGATGACAAA..................... 627
         ::   |||         :::::::::   :::
      74 erSerHisLeuThrLeuTrpSerGluGlyGlnGlyArgGlyGlyValAsn 90

     628 ..........................CACGGCGATTTATCC...TACTC 647
                                  :::|||:::|||:::   |||||
      91 ValThrGlyAsnIleThrSerThrThrAsnGlyAsnLeuThrIleTyrSe 107

     648 CGGCGCATGGTTA...ATTGGCGGCAATACACATATGCAGGGTTGGGGAA 694
         ||||:::|||:::   :::      |||      :::::::::   |||
     107 rGlyGlyTrpValAspValHisLysAsnIleThrLeuLysSer...GlyT 123

     695 ATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTAT 744
               ::::::::::::   :::||||||:::          :::
     123 yrLeuAsnIleThrThrLysGlnGlyAspIleAlaPheGluAspLysPro 139

     745 GGC............CCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGG 782
         |||              :::        |||:::   :::   :::::
     140 GlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaValAsnAs 156

     783 TTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACG 832
         ::::      |||   :::|||:::::::::            |||||||
     156 nLysLysGlyPheArgPheAspAsnValThr............LeuAsnG 169

     833 GAGTT.......................TTACAAACCGGCTACCCTTAT 858
         ||:::                       ::::::||||||
     169 lyThrGlyGlyGlyLeuSerPheLysTyrIleGluThrGlyAsnArgAsp 185

     859 TCCGGCAGGGAAAACGGTTTC........................... 879
         |||:::   |||:::   |||
     186 SerAsnPheGluThrHisPheArgGlyArgLeuAsnIleSerGlyLysVa 202

     880 .............CAGCTGATACGCAAAGATTGGTTCTACGATGACATTT 916
                     |||      :::::::::|||
     202 lAspIleLeuMetGlnAlaArgGlnGluAsnTrpAsnArgArgHisTrpG 219

     917 ACAGAGGCGATACACATACCGTCTNTTTTGAACCGCGCAGTAACGGACAT 966
           |||:::      :::::::::   ::::::      :::|||::::::
     219 lyArgSerHisTrpAsnValThrArgLeuAsnValSerGluAsnSerTyr 235

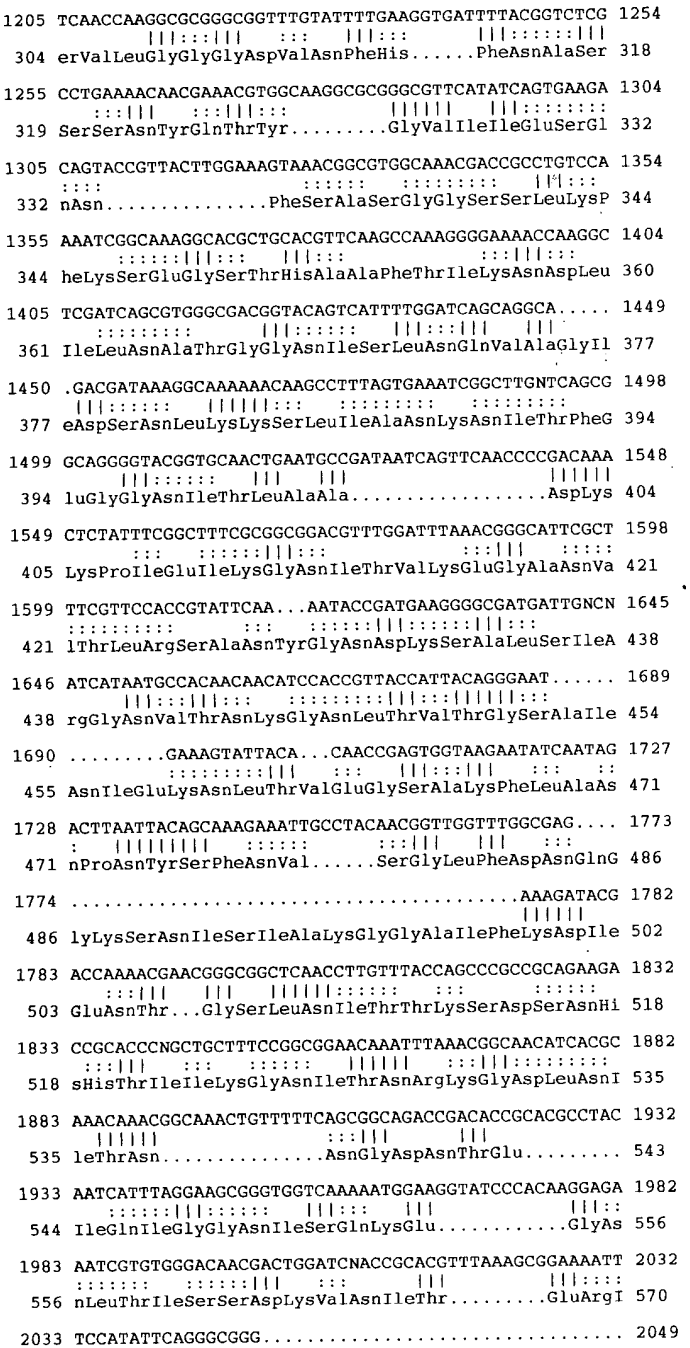
     967 TTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAA 1016
         |||:::   |||   :::::::::|||:::
     236 PheAsnValThrIleAspSerSerGlySer..................... 245

    1017 CGAAAAGGTNTCCAATCCA.........AAGCTTAAAGTACAGACAGTCC 1057
               :::|||:::|||            |||::::::|||:::
     246 .......AlaSerSerProGlyAlaGlyProLeuAsnAlaGlnSer...G 259

    1058 GACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCG 1107
           |||         |||:::|||:::                 :::   |||
     259 lyLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnIleAlaAla 275

    1108 GCAGGGGGTGTTAAT...CAGTACCGTCCAAGGTTAAACAACGGTGAAAA 1154
         :::::::::||||||   :::   :::|||   :::
     276 SerSerAlaValAsnPheAsnIleLysProProIle.............. 287

    1155 CCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACA 1204
         ::::::   :::      ||||||      |||:::   :::|||
     288 .ValAspLysValThrAsnGlyAsnHisThrLeuPheLysGlyAsnIleS 304
```


1205 TCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCG 1254
304 erValLeuGlyGlyGlyAspValAsnPheHis.....PheAsnAlaSer 318
1255 CCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGA 1304
319 SerSerAsnTyrGlnThrTyr.....GlyValIleIleGluSerGl 332
1305 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1354
332 nAsn.....PheSerAlaSerGlyGlySerSerLeuLysP 344
1355 AAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGC 1404
344 heLysSerGluGlySerThrHisAlaAlaPheThrIleLysAsnAspLeu 360
1405 TCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA..... 1449
361 IleLeuAsnAlaThrGlyGlyAsnIleSerLeuAsnGlnValAlaGlyIl 377
1450 .GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCG 1498
377 eAspSerAsnLeuLysLysSerLeuIleAlaAsnLysAsnIleThrPheG 394
1499 GCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAA 1548
394 luGlyGlyAsnIleThrLeuAlaAla.....AspLys 404
1549 CTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCT 1598
405 LysProIleGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnVa 421
1599 TTCGTTCCACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGNCN 1645
421 lThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 438
1646 ATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGAAT..... 1689
438 rgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 454
1690 GAAAGTATTACA...CAACCGAGTGGTAAGAATATCAATAG 1727
455 AsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAs 471
1728 ACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG.... 1773
471 nProAsnTyrSerPheAsnVal.....SerGlyLeuPheAspAsnGlnG 486
1774 AAAGATACG 1782
486 lyLysSerAsnIleSerIleAlaLysGlyGlyAlaIlePheLysAspIle 502
1783 ACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGA 1832
503 GluAsnThr...GlySerLeuAsnIleThrThrLysSerAspSerAsnHi 518
1833 CCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGC 1882
518 sHisThrIleIleLysGlyAsnIleThrAsnArgLysGlyAspLeuAsnI 535
1883 AAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTAC 1932
535 leThrAsn.....AsnGlyAspAsnThrGlu..... 543
1933 AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGA 1982
544 IleGlnIleGlyGlyAsnIleSerGlnLysGlu.....GlyAs 556
1983 AATCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATT 2032
556 nLeuThrIleSerSerAspLysValAsnIleThr.....GluArgI 570
2033 TCCATATTCAGGGCGGG..... 2049
570 leThrIleLysAlaGlyValAsnGlyAspAsnSerAspSerAsnGluAla 586
2050 CAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGA 2090
587 ThrSerAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAs 603
2091 TTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATC 2140
603 pLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspAsnS 620
2141 AAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGT 2190
620 erAsnLeuThrIleGlyAspAsnSerAsp...AlaGlyAsnThrAspAla 635
2191 GTCGAA.....NAAANCATTACCGACGATAAAGTGATTGCTTCA.. 2229
636 LysLysValThrPheSerAsnValLysAspSerLysIleSerAlaSerAs 652
2230 TTGACTAAGACNGACNTNAGCGGCANTGTNAGNC 2263
652 pHisAsnValThrLeuAsnSerLysValGluThrSerGlyAspThrAspS 669
2264 TNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC 2313
669 erThrGluAspGlyGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLys 685
2314 AATCTTAGTGCAAATGGCGAT.....ACACGTTATAC 2345
686 AsnValThrValAsnAsnAsnIleThrSerHisLysThrValAsnIleTh 702
2346 AGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC.....G 2386
702 rAlaSerGluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrT 719
2387 TGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC... 2433
719 hrGlySerValGluValThrAlaLysThrGlyAspIleLysGlyGlyIle 735
2434 NCATCGGNTTCGGGCAATGCT.....TCATTTAA 2462
736 GluSerAsnSerGlyAsnValAsnIleThrAlaSerGlyAspThrLeuAs 752
2463 TCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACG 2512
752 nValSerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyA 769
2513 CTAAGGCAAACGTAAGCCATTCCGCACTCAAC.....GGCAATGTC 2553
769 laValThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAla 785
2554 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGG 2603
786 AsnIleThrThrLys.....ThrGl 792
2604 ACAACTCAGCGGCAGCAAGGANACAGCA.....TTACACTTAAAAG 2644
792 yGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrA 809
2645 ACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTT 2694
809 laSerGlyAsnThrLeu.....AsnValSerAsnIleThrGly 821
2695 GACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGG 2744
822 GlnAsnValThrValThrAlaAsnSer..... 830
2745 CGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTT 2794
831 GlyAlaIleThrThrThrGluGly..... 838
2795 CCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2844

```
 839 ........SerThrIleAsnAlaThrThrGlyAspAlaAsnIleThrThr 852

2845 CTGACG......GTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTT 2888
       |||       :::|||||||||:::
 853 GlnThrGlyAsnIleAsnGlyLysVal..................... 861

2889 TATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAA 2938
                                                   ||||
 862 ..............................................GluS 863

2939 GTTCCGAAGGNACT..................TACACCTTGGCGGTC 2967
     |||||:::|||:::                  ||||||||||||
 863 erSerSerGlySerValThrLeuIleAlaThrGlyGlnThrLeuAlaVal 879

2968 AACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGA 3017
     :::|||   :::::::::   |||:::        :::|||:::   ::
 880 GlyAsnIleSerGlyAspThrValThr.........IleThrAlaAspLy 893

3018 AGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAA 3067
     :||||||   :::          |||:::::::::|||
 893 sGlyLysLeuThrThrGlnThrSerSerLysIleAsn.............. 905

3068 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3117
                   |||:::                        ::::::
 906 .................GlyThrLysSerValThrThrSerSerGlnSer 916

3118 GGCGAGTTCCGC.........CTGCATAATCCGGTCAAAGAACAAGAGCT 3158
     |||::::::                  |||   |||:::   :::
 917 GlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaTh 933

3159 TTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACA 3208
      ::::::         :::|||   :::|||   :::|||:::     :
 933 rGlySerLeuThrThrGlnAlaGlySerLysIleGluAlaLysThrGlyG 950

3209 ACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAA 3258
     ::|||:::      :::|||      :::   |||        :::
 950 luAlaAsnValThrSerAlaThrGlyThrIleGlyGlyThrIleSerGly 966

3259 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGT 3308
     :::|||   :::|||               :::|||:::   :::|||::
 967 AsnThrValAsnVal..............ThrAlaAsnThrAspAsnLe 978

3309 CGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAG 3358
     :   |||                                      ||||
 978 uThrIle.....................................LysA 982

3359 ACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACC 3408
     ||:::|||      |||         :::   :::      ::::::|||
 982 spGlyAlaArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaThr 998

3409 GCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCC 3458
     :::       :::   :::      |||:::   :::   :::
 999 GlyGlyThrLeuThrThrGluThrSerSerAspIleThrSerSerAsnGl 1015

3459 CCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATG 3508
      |||    :::|||:::   :::   |||
1015 yGln.ThrThrLeuThrAlaLysAspSer..................... 1024

3509 CCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC 3558
                                     |||   |||   |||
1025 ................................SerIleAlaGlySerIl 1030

3559 GTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGT 3608
           :::|||   |||                |||   :::|||
1030 eAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValA 1047

3609 TTGGA.........CAAGCNGCATCCGGNACACCAAACACTACCGTTCGC 3649
      |||          :::|||||||||||||||
1047 laGlySerLysIleGluAlaAlaSerGlyThr..............Leu 1058

3650 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA........ 3691
       |||:::|||      |||:::          :::|||:::
1059 ValIleAsnAlaLysAspAlaGlnLeuAspGlyAlaAlaSerGlyAspHi 1075

3692 .TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTC 3740
      :::|||   :::   |||:::|||   |||   |||:::|||   :::
1075 sThrValValAsnAlaThrAsnAlaAsnGly...SerGlySerValIleA 1091

3741 GCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCAC 3790
       ||||||:::   ::::::      |||            |||   :::
1091 laThrThrSerSerArgValAsnIleThrGlyAspLeuIleThrIleAsn 1107

3791 GGCTTGCCCACGGCGCCGTTTTCGGGC...AATACGGCATCGGCAGGTTC 3837
     ||||||               :::|||   ||||||:::         ::
1108 GlyLeuAsnIleIleSerLysAsnGlyLysAsnThrValLeuLeuLysGl 1124

3838 GACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGA 3887
     :::::::   :::::::::           :::   |||:::   :::
1124 yValGluIleAspValLysTyrIleGlnProGlyIleAlaSerValAsnG 1141

3888 CGGCATCGGAGGCAAAATCCGCC 3910
       :::   |||||||||   |||
1141 luValIleGluAlaLysArgAla 1148
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828

seq_documentation_block:
ID   AAB01828 standard; Protein; 1228 AA.
XX
AC   AAB01828;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K1.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 313
FT                   /note= "Encoded by GG"
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52179.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 20A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
```

```
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1228 AA;


alignment_scores:
            Quality:  264.00            Length:   1325
              Ratio:  0.401               Gaps:     60
 Percent Similarity:  49.736   Percent Identity: 20.377

alignment_block:
 US-09-303-518D-651 x AAB01828   ..

 Align seg 1/1  to: AAB01828  from: 1  to: 1228

     346 AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGA 395
         |||   :::   |||         :::|||         ::::::|||::
       9 AsnValSerIleAsnAlaProAlaLeuGlyArgThrGluSerThrProAs  25

     396 TCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTG 445
         :::::::      :::|||:::      ::::::   |||||||||
      25 nAsnAsn......GluTyrAspSerProAsnGlnIleAsnTyrLysAsnL  40

     446 ACAATTCACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAA 495
           |||                 ::::::   :::   |||:::   |||
      40 ysProSerLeuSerThrLeuThrAsnThrThrLeuGluArgIleLeuLys  56

     496 TTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAA 545
           |||:::         |||::::::|||:::             ::
      57 ArgAsnThrSer.........ValAsnIleThrAlaThrLysThrIleTh  70

     546 TACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAG 595
         ::::   ||||||                    ::::::||||||::::
      70 rValAsnSerAsp.....................IleAsnIleGlyAspS  80

     596 GACACCACTATTGGCGTTATGATGATGACAAA..................  627
         ::   |||         :::::::::   :::
      80 erSerHisLeuThrLeuTrpSerGluGlyGlnGlyArgGlyGlyValAsn  96

     628 ..........................CACGGCGATTTATCC...TACTC  647
                                  :::|||:::|||:::   |||||
      97 ValThrGlyAsnIleThrSerThrThrAsnGlyAsnLeuThrIleTyrSe 113

     648 CGGCGCATGGTTA...ATTGGCGGCAATACACATATGCAGGGTTGGGGAA 694
         ||||:::|||:::   :::      |||      :::::::::   |||
     113 rGlyGlyTrpValAspValHisLysAsnIleThrLeuLysSer...GlyT 129

     695 ATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTAT 744
               ::::::::::::   :::|||||||:::          :::
     129 yrLeuAsnIleThrThrLysGlnGlyAspIleAlaPheGluAspLysPro 145

     745 GGC............CCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGG 782
         |||              :::         |||:::   :::   :::::
     146 GlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaValAsnAs 162

     783 TTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACG 832
         ::::      |||   :::|||:::::::::          |||||||
     162 nLysLysGlyPheArgPheAspAsnValThr............LeuAsnG 175

     833 GAGTT.......................TTACAAACCGGCTACCCTTAT 858
         ||:::                      ::::::||||||
     175 lyThrGlyGlyGlyLeuSerPheLysTyrIleGluThrGlyAsnArgAsp 191

     859 TCCGGCAGGGAAAACGGTTTC.......................... 879
         |||:::   |||:::   |||
     192 SerAsnPheGluThrHisPheArgGlyArgLeuAsnIleSerGlyLysVa 208

     880 .............CAGCTGATACGCAAAGATTGGTTCTACGATGACATTT 916
                     |||       ::::::::::|||
     208 lAspIleLeuMetGlnAlaArgGlnGluAsnTrpAsnArgArgHisTrpG 225

     917 ACAGAGGCGATACACATACCGTCTNTTTTGAACCGCGCAGTAACGGACAT 966
           |||:::      :::::::::   ::::::      :::|||::::::
     225 lyArgSerHisTrpAsnValThrArgLeuAsnValSerGluAsnSerTyr 241

     967 TTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAA 1016
         |||:::   |||   :::::::::|||:::
     242 PheAsnValThrIleAspSerSerGlySer..................... 251

    1017 CGAAAAGGTNTCCAATCCA.........AAGCTTAAAGTACAGACAGTCC 1057
                ::::|||:::|||            |||:::::::|||:::
     252 .......AlaSerSerProGlyAlaGlyProLeuAsnAlaGlnSer...G 265

    1058 GACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCG 1107
           |||         |||:::|||:::                  :::   |||
     265 lyLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnIleAlaAla 281

    1108 GCAGGGGGTGTTAAT...CAGTACCGTCCAAGGTTAAACAACGGTGAAAA 1154
         :::::::::||||||   :::   :::|||   :::
     282 SerSerAlaValAsnPheAsnIleLysProProIle.............. 293

    1155 CCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACA 1204
          ::::::   :::       ||||||      |||:::   ::::||
     294 .ValAspLysValThrAsnGlyAsnHisThrLeuPheLysGlyAsnIleS 310

    1205 TCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCG 1254
               |||:::|||   :::   |||:::      |||:::::|||
     310 erValLeuGlyGlyGlyAspValAsnPheHis......PheAsnAlaSer 324

    1255 CCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGA 1304
         :::|||   :::|||:::          ||||||   |||::::::::
     325 SerSerAsnTyrGlnThrTyr.........GlyValIleIleGluSerGl 338

    1305 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1354
         ::::                  ::::::   :::::::::   |||:::
     338 nAsn...............PheSerAlaSerGlyGlySerSerLeuLysP 350

    1355 AAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGC 1404
               ::::::|||:::   |||:::            :::|||:::
     350 heLysSerGluGlySerThrHisAlaAlaPheThrIleLysAsnAspLeu 366

    1405 TCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA..... 1449
           :::::::::      |||::::::   |||:::|||   |||
     367 IleLeuAsnAlaThrGlyGlyAsnIleSerLeuAsnGlnValAlaGlyIl 383

    1450 .GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCG 1498
          |||::::::   ||||||:::   :::::::::   :::::::::
     383 eAspSerAsnLeuLysLysSerLeuIleAlaAsnLysAsnIleThrPheG 400

    1499 GCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAA 1548
           |||::::::   |||   |||                  ||||||
     400 luGlyGlyAsnIleThrLeuAlaAla.................AspLys 410

    1549 CTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCT 1598
           :::   ::::::|||:::                  :::|||   :::::
     411 LysProIleGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnVa 427

    1599 TTCGTTCCACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGNCN 1645
```

```
 427 lThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 444

1646 ATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGAAT...... 1689
 444 rgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 460

1690 .........GAAAGTATTACA...CAACCGAGTGGTAAGAATATCAATAG 1727
 461 AsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAs 477

1728 ACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG.... 1773
 477 nProAsnTyrSerPheAsnVal......SerGlyLeuPheAspAsnGlnG 492

1774 .....................................AAAGATACG 1782
 492 lyLysSerAsnIleSerIleAlaLysGlyGlyAlaIlePheLysAspIle 508

1783 ACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGA 1832
 509 GluAsnThr...GlySerLeuAsnIleThrThrLysSerAspSerAsnHi 524

1833 CCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGC 1882
 524 sHisThrIleIleLysGlyAsnIleThrAsnArgLysGlyAspLeuAsnI 541

1883 AAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTAC 1932
 541 leThrAsn..............AsnGlyAspAsnThrGlu......... 549

1933 AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGA 1982
 550 IleGlnIleGlyGlyAsnIleSerGlnLysGlu............GlyAs 562

1983 AATCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATT 2032
 562 nLeuThrIleSerSerAspLysValAsnIleThr.........GluArgI 576

2033 TCCATATTCAGGGCGGG.............................. 2049
 576 leThrIleLysAlaGlyValAsnGlyAspAsnSerAspSerAsnGluAla 592

2050 .........CAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGA 2090
 593 ThrSerAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAs 609

2091 TTGNCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCATC 2140
 609 pLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspAsnS 626

2141 AAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGT 2190
 626 erAsnLeuThrIleGlyAspAsnSerAsp...AlaGlyAsnThrAspAla 641

2191 GTCGAA.........NAAANCATTACCGACGATAAAGTGATTGCTTCA.. 2229
 642 LysLysValThrPheSerAsnValLysAspSerLysIleSerAlaSerAs 658

2230 ...............TTGACTAAGACNGACNTNAGCGGCANTGTNAGNC 2263
 658 pHisAsnValThrLeuAsnSerLysValGluThrSerGlyAspThrAspS 675

2264 TNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC 2313
 675 erThrGluAspGlyGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLys 691

2314 AATCTTAGTGCAAATGGCGAT.................ACACGTTATAC 2345
 692 AsnValThrValAsnAsnAsnIleThrSerHisLysThrValAsnIleTh 708

2346 AGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC.........G 2386
```

```
 708 rAlaSerGluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrT 725

2387 TGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC... 2433
 725 hrGlySerValGluValThrAlaLysThrGlyAspIleLysGlyGlyIle 741

2434 NCATCGGNTTCGGGCAATGCT.....................TCATTTAA 2462
 742 GluSerAsnSerGlyAsnValAsnIleThrAlaSerGlyAspThrLeuAs 758

2463 TCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACG 2512
 758 nValSerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyA 775

2513 CTAAGGCAAACGTAAGCCATTCCGCACTCAAC.........GGCAATGTC 2553
 775 laValThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAla 791

2554 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGG 2603
 792 AsnIleThrThrLys............................ThrGl 798

2604 ACAACTCAGCGGCAGCAAGGANACAGCA.........TTACACTTAAAAG 2644
 798 yGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrA 815

2645 ACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTT 2694
 815 laSerGlyAsnThrLeu............AsnValSerAsnIleThrGly 827

2695 GACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGG 2744
 828 GlnAsnValThrValThrAlaAsnSer..................... 836

2745 CGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTT 2794
 837 .........GlyAlaIleThrThrThrGluGly................ 844

2795 CCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2844
 845 ........SerThrIleAsnAlaThrThrGlyAspAlaAsnIleThrThr 858

2845 CTGACG......GTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTT 2888
 859 GlnThrGlyAsnIleAsnGlyLysVal..................... 867

2889 TATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAA 2938
 868 .............................................GluS 869

2939 GTTCCGAAGGNACT.....................TACACCTTGGCGGTC 2967
 869 erSerSerGlySerValThrLeuIleAlaThrGlyGlnThrLeuAlaVal 885

2968 AACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGA 3017
 886 GlyAsnIleSerGlyAspThrValThr.........IleThrAlaAspLy 899

3018 AGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAA 3067
 899 sGlyLysLeuThrThrGlnThrSerSerLysIleAsn............. 911

3068 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3117
 912 ................GlyThrLysSerValThrThrSerSerGlnSer 922

3118 GGCGAGTTCCGC.........CTGCATAATCCGGTCAAAGAACAAGAGCT 3158
 923 GlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaTh 939

3159 TTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACA 3208
 939 rGlySerLeuThrThrGlnAlaGlySerLysIleGluAlaLysThrGlyG 956
```

```
  3209 ACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAA 3258
       ::|||:::      :::|||       :::   |||       :::
   956 luAlaAsnValThrSerAlaThrGlyThrIleGlyGlyThrIleSerGly 972

  3259 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGT 3308
       :::|||   :::|||              :::|||:::   :::|||::
   973 AsnThrValAsnVal.............ThrAlaAsnThrAspAsnLe 984

  3309 CGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAG 3358
       :   |||                                   ||||
   984 uThrIle.....................................LysA 988

  3359 ACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACC 3408
       ||:::|||      |||         :::   :::      ::::::|||
   988 spGlyAlaArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaThr 1004

  3409 GCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCC 3458
       :::       :::      :::      |||:::   :::   :::
  1005 GlyGlyThrLeuThrThrGluThrSerSerAspIleThrSerSerAsnGl 1021

  3459 CCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATG 3508
        |||    :::|||:::   :::   |||
  1021 yGln.ThrThrLeuThrAlaLysAspSer..................... 1030

  3509 CCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC 3558
                                     |||   |||   |||
  1031 ..............................SerIleAlaGlySerIl 1036

  3559 GTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGT 3608
             :::|||   |||              |||   :::|||
  1036 eAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValA 1053

  3609 TTGGA.........CAAGCNGCATCCGGNACACCAAACACTACCGTTCGC 3649
         |||         :::|||||||||||||||
  1053 laGlySerLysIleGluAlaAlaSerGlyThr.............Leu 1064

  3650 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA........ 3691
          |||:::|||      |||:::         :::|||:::
  1065 ValIleAsnAlaLysAspAlaGlnLeuAspGlyAlaAlaSerGlyAspHi 1081

  3692 .TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTC 3740
        :::|||   :::   |||:::|||   |||   |||:::|||   :::
  1081 sThrValValAsnAlaThrAsnAlaAsnGly...SerGlySerValIleA 1097

  3741 GCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCAC 3790
         ||||||:::   ::::::      |||                |||  :::
  1097 laThrThrSerSerArgValAsnIleThrGlyAspLeuIleThrIleAsn 1113

  3791 GGCTTGCCCACGGCGCCGTTTTCGGGC...AATACGGCATCGGCAGGTTC 3837
       ||||||               :::|||   ||||||:::         ::
  1114 GlyLeuAsnIleIleSerLysAsnGlyLysAsnThrValLeuLeuLysGl 1130

  3838 GACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGA 3887
       :::::::   :::::::::             :::   |||:::   :::
  1130 yValGluIleAspValLysTyrIleGlnProGlyIleAlaSerValAsnG 1147

  3888 CGGCATCGGAGGCAAAATCCGCC 3910
         :::   |||||||||   |||
  1147 luValIleGluAlaLysArgAla 1154
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825

seq_documentation_block:
ID AAB01825 standard; Protein; 1221 AA.
XX
AC AAB01825;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain Joyc mature HMW1A protein, SEQ ID NO:28.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain Joyc.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52176.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 18A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1221 AA;

alignment_scores:
Quality: 263.50 Length: 1160
Ratio: 0.473 Gaps: 62
Percent Similarity: 48.017 Percent Identity: 21.121

alignment_block:
US-09-303-518D-651 x AAB01825 ..

Align seg 1/1 to: AAB01825 from: 1 to: 1221

```
   391 CCCGATCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
       ||||||:::      ::::::           :::|||||||:::   ::
     1 ProAspAsnValSerIleAsnAlaGlyThrSerGluArgAsnAspAlaSe 17
```

```
441 GCCTGACAATTCACACCCTTACAAC...GGCGATTANCATATGCCGCGT. 486
    :|||  ::::::  |||  :::  |||          .|||:::
 17 rProThrGluAspPheProThrGlyAlaGlyGlyLysAspAsnProLysL 34

487 .....TTGCATAAATTT 498
                    |||::::::::::
 34 ysAsnAlaHisAsnLysProThrLeuIleAsnThrThrLeuGluArgIle 50

499 GTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATAC 548
    ::::::          |||:::::|||:::     |||   :::::
 51 LeuSerGlyAsnThrPheValAsnIleThrAlaArgLysArgIleThrVa 67

549 CTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGAC 598
    :   ||||||                 ::::::|||  ::::::
 67 lAsnSerAsp.....IleAsnIleLysAspSerS 77

599 ACCAC......TATTGGCGTTATGATGATGACAAACACGGC...GATTTA 639
     |||       |||     :::|||::::::   |||   |||:::
 77 erHisLeuIleLeuTrpSerGluAsnAspAsnSerSerGlyValAspIle 93

640 TCC.....TACTCCGG 650
    :::          ||||||::
 94 LysGlyAsnIleThrSerThrThrGlyGlySerLeuThrIleTyrSerSe 110

651 CGCATGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAATG 700
    :::|||:::        :::   |||  :::        |||:::|
110 rGlyTrpIle.....AspIleHisLysAsnIleThrLeuAsnSerG 124

701 GCGTANTTAGTTTGAGC.....GGCGATGTGCGCCATGCCAACGAC 741
    ||:::   :::::::::       ||||||:::        :::
124 lyLeuLeuAsnIleThrThrLysGlnGlyAspIleAlaPheGluLysGly 140

742 TATGGCCCTATGCCGATTGCAGGT.....GCGGCAGGCGACAG 779
      :::|||     |||:::|||          :::||||||:::::
141 AsnAsnPro...ThrIleThrGlyGlnGlyThrIleThrAlaGlyAsnGl 156

780 C.....GGTTCGCCAA 790
    :                 |||:::  :
156 yLysGlyPheArgPheGluAsnAlaSerLeuAsnGlyIleGlyThrGlyL 173

791 TGTTTATT.....TATGACAAAACAAACAATAAATGGCTGCTC... 828
    ::::::::          |||     ||||||     ::::::
173 euLeuPheAsnIleLysArgAspLeuGlyAsnAsnPheGlnIleIleAsn 189

829 ......AACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAA 872
          ||||||:::|||:::              ||||||:::  ||
190 PhePheAsnGlyThrLeuAsn.....IleSerGlyLysValAs 202

873 CGGTTTCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAG 922
    |      :::::::|||   |||  |||  |||:::      :::||||
202 nIleSerMetValIleProLysLysTrpAspTyrSerLys...PheArgG 218

923 GCGATACACAT......ACCGTCTNTTTTGAACCGCGCAGTAACGGACAT 966
    ||   |||:::      ::::::  ::::::      :::::::::
218 lyArgThrTyrTrpAsnValThrHisLeuAsnValSerGluGlySerLys 234

967 TTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAA 1016
    |||::::::|||   :::::::::|||      |||:::        ||
235 PheAsnLeuThrIleAspSerArgGlyAspAspThrAlaGlyThrLeuAs 251

1017 CGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTG 1066
     |       :::       ::::::  :::        :::|||||
251 nThrProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrIlePheA 268

1067 ACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGT 1116
     ||  :::  |||  :::           :::  |||      |||
268 spValLysGlnAsnGlyAlaValThrPheAspIleLysAlaProIleGly 284

1117 GTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTAT 1166
     |||                    ||||||:::::::||||||::::::
285 Val.....AsnAsnAsnArgAsnLeuAsnTyrAl 294

1167 CGATTACGGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCG 1216
```

```
       ::::::  ||||||::::::  :::|||              |||:
294 aSerPhe...AsnGlyAsnIleSerValSer.....GlyG 305

1217 CGGGCGGTTTGTATTTTGAA..... 1236
     ::|||::::::  |||:::
305 lyGlyAsnValThrPheLysLeuLeuAlaSerSerSerThrAlaGlnThr 321

1237 ...GGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGC 1283
        |||  |||      |||  :::  |||  :::      :::::
322 ProGlyValPheIleAsnSerLysHisPheAsnAlaSerGlyGlySerSe 338

1284 GGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCG 1333
     :          :::|||  ||||||        ||||||:::   :
338 rLeuGluPheArgThrGluGlySerThr.....LysValGlyPheL 352

1334 TGGCAAACGACCGCCTGTCC..... 1353
     ::   |||:::  |||:::
352 euIleAsnAsnAspLeuThrLeuAsnAlaThrGlyGlyAsnIleSerLeu 368

1354 .....AAAATCGGCAAAGGCACGCTGCACGT 1379
                              ||||||||||||      ||
369 LeuGlnValGluGlyIleAspGlyMetIleGlyLysGly.....Va 382

1380 TCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAG 1429
     |   ||||||                 :::|||:::   :::   |||::::
382 lValAlaLys.....LysAsnIleThrPheAlaGlyGlyAsnI 395

1430 TCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGT 1479
     ::  :::                    |||:::|||:::|||::::::
395 leThrPhe.....GlySerLysLysAlaIleThr 404

1480 GAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCC..... 1524
     ||||||        :::|||::::::|||::::::   ||||||
405 GluIle.....GluGlyAsnAlaThrIleAsnAsnAsnAlaAsnVa 418

1525 .....GATAATCAGTTCAACCCC..... 1542
                          ||||||:::   :::|||
418 lThrLeuIleGlySerAspPheAspAsnHisGlnLysProLeuThrIleL 435

1543 .....GACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAAC 1587
          |||  :::         :::::::::|||:::|||      :::
435 ysLysAspValIle.....IleAsnSerGlyAsnLeuThrAlaGly 448

1588 GGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGAT 1637
     |||:::   ::::::::::::        ::::::::::::||||||
449 GlyAsnValIleAsnIleAsnGlyAsnLeuThrValAsnAsnGlyAlaAs 465

1638 GATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATT..... 1680
      :::         :::|||   |||:::   |||   ::::::
465 nLeu.....LysAlaIleThrAsnPheThrPheAsnValGlyGlyL 479

1681 ..........ACAGGGAATGAAAGTATTACACAACCGAGTGGT..... 1713
                    ||||||::::::|||:::         |||
479 euPheAspAsnLysGlyAsnSerAsnIleSerIleAlaArgGlyGlyAla 495

1714 ......AAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAA 1757
           |||:::||||||::   :::   |||   :::|||:::   ||
496 LysPheLysAspIleAsnAsnThrSer...SerLeuAsnIleThrThrAs 511

1758 CGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACC 1807
     |                ::::|||||||||
511 n.....SerAspThrThr..... 515

1808 TTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGA 1857
        |||:::          |||  :::  :::      :::|||
516 .....TyrArgThrIleIleGluGlyAsnIleThrAsnLysAlaGlyAsp 530

1858 ACAAAT.....TTAAACGGCAA 1874
       |||                  ::::::|||||
531 LeuAsnIleIleAspAsnLysGlyAsnAlaGluIleGlnIleGlyGlyAs 547

1875 CATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGA...CAC 1921
     ||||:::|||   :::|||:::|||   :::||| :::|||:::   :::
```

```
 547 nIleSerGlnLysGluGlyAsnLeuThrIleSer.SerAspLysIleAsn 563

1922 CGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATC 1971
      |||      |||      :::   |||
 564 IleThrAsnGlnIleThrIleLysLysGly.................Va 574

1972 CCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGC....ACGT 2017
     :::||||||   |||   :::::::|||:::::::::::::      |||:
 574 lAsnLysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLeuThrI 591

2018 TTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGC 2067
     ::|||:::::::::::::::::   |||         :::   :::
 591 leLysThrLysGluLeuGlnLeuThrGlyAspLeuAsnIleSerGlyPhe 607

2068 AATGTTGCCAAAGTGGAA.........................GGCGA 2090
     :::   |||::::::                            |||::
 608 AspLysAlaGluIleThrAlaLysGluGlyAlaAspLeuIleIleGlyAs 624

2091 TTGNCATTTGAGCAATCACGCCCAAGCA.........GTTTTTGGTGTCG 2131
     ::::      :::|||:::|||::|||      :::|||:::   :
 624 nSerAspAsnAsnAsnAsnAlaAsnAlaLysLysValThrPheAsnGlnV 641

2132 CACCGCATCAA..............AGCCATACAATCTGTACACGTTCG 2166
     ::      :::                |||||:::::      :::|||
 641 alLysAspSerLysIleSerAlaGlySerHisAsnValThrLeuAsnSer 657

2167 ..............................................GA 2168
                                                    ||
 658 LysValGluThrSerAsnGlyAsnAsnAspAlaGluSerAsnAsnGlyAs 674

2169 CTGGACNGGTCTGACA..............AATTGTGTCGAANAAANCA 2203
     |   |||:::||||||                  |||:::      |
 674 pSerThrSerLeuThrIleAsnAlaLysAsnValThrValAsnAsnAsnI 691

2204 TTACCGACGATAAAGTGATT............................. 2223
     ||||||::   |||::::::                            .
 691 leThrSerHisLysThrValAsnIleThrAlaSerGluAsnValThrThr 707

2224 ...GCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNA 2270
        |||:::   |||   :::   :::   |||:::|||   ::::::::
 708 LysAlaGlyThrThrIleAsnAlaThrIleGlySerValGluValThrAl 724

2271 TNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC......A 2314
     :   :::                  |||      :::   :::|||     |
 724 aLysThrGlyAspIleLysGlyGlyIleGluSerAsnSerGlyAsnValA 741

2315 ATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC 2364
     ||:::::|||:::|||||||||      :::||||||:::   :::
 741 snIleThrAlaSerGlyAspThr...LeuAsnValSerAsnIleThrGly 756

2365 CAAAACGGCAACCTT.............................. 2379
     ||||||   ::::::                               .
 757 GlnAsnValThrValAlaAlaAlaSerGlyAlaValThrThrThrLysGl 773

2380 .............AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAG 2416
                  :::   :::||||||||::   |||   :::   :
 773 ySerThrIleAsnAlaThrThrGlyAsnAlaAsnIleThrThrLysThrG 790

2417 CCACATTAAACGGC.........................AACNCATCG 2439
     ::   :::||||||                        |||   :::
 790 lyGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThr 806

2440 GNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGG 2489
     :::||||||||::      |||:::||||||   ::::::||||||
 807 AlaSerGlyAsnThrLeuAsnAsnValSerAsnIleThrGlyGlnAsnVa 823

2490 CAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCAC 2539
      ::::::|||   ::::::::::|||   ::::::::::::   |||::::
 823 lThrValThrAlaAsnSerGlyAlaIleThrThrThrGluGlySerThrI 840

2540 TCAAC.........GGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT 2580
     ::|||         |||:::::::::::::::   :::
 840 leAsnAlaThrThrGlyAspAlaAsnIleThrThrGln............ 852
```

```
2581 TTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGC 2630
                          ||||||::::::::::|||:::      :::::
 853 ..................ThrGlyAsnIleAsnGlyLysValGluSerSe 863

2631 ATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA. 2679
     :         :::   |||         :::   :::|||      |||
 863 r.........SerGlySerValThrLeuIleAlaThrGlyGlnThrLeuA 877

2680 .....GGCAATTTAAACCTTGACAACGCCACCATTACACTCAAT...... 2718
          ||||||::::::   |||:::::::|||||||||   :::
 877 laValGlyAsnIleSerGlyAspThrValThrIleThrAlaAspLysGly 893

2719 ..............................TCCGCCTATCGCCACGATGC 2738
                                   ::::::   :::      ::
 894 LysLeuThrThrGlnThrSerSerLysIleAsnGlyThrLysSerValTh 910

2739 TGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGC 2788
     ::::::::::|||:::|||   :::|||   |||
 910 rThrSerSerGlnSerGlyAspIleSerGlyThr................ 921

2789 GCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTC 2838
                                       ::::::::::::
 922 ................................IleSerGlyAsnThr 926

2839 AACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTT 2888
        ::::::::::::::::|||:::|||:::::::|||:::
 927 ValSerValSerAlaThrGlySerLeuThrThrGlnAla........... 939

2889 TATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAA 2938
                                  ::::::|||::::::   |||:::::
 940 ........................GlySerLysIleGlu...AlaLysT 947

2939 GTTCCGAAGGNACTTACACCTTGGCGGTC..................... 2967
     ::::::|||::::::   |||   |||:::
 947 hrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGlyThrIle 963

2968 ...AACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGT 3014
        :::|||||   |||      ::::::   |||:::|||||||:::
 964 SerGlyAsnThrValAsnValThrAlaAsnThrAspAsnLeuThrIleLy 980

3015 GGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTC....... 3057
      :::|||      |||      :::      :::::::::
 980 sAspGlyAlaArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaT 997

3058 ........ACCCTGCAAAACGAACACGTCG.........ATGCCGGCGCG 3090
             ||||||   |||:::||||||                 :::::
 997 hrGlyGlyThrLeu.ThrThrGluThrSerSerAspIleThrSerSerAs 1013

3091 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCC 3140
     :|||   |||:::   :::||||||   :::|||   |||      |||:
1013 nGlyGlnThrThrLeuThrAlaLysAspSerSerIleAlaGlySerIleA 1030

3141 GGTCAAAGAACAAGAGCTTTCCGACA 3166
     :::::   |||   ::::::   |||
1030 snAlaAlaAsnValThrLeuAsnThr 1038
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826

seq_documentation_block:
ID AAB01826 standard; Protein; 975 AA.
XX
AC AAB01826;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain Joyc HMW2A protein, SEQ ID NO:30.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.

XX
OS Haemophilus influenzae strain Joyc.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52177.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 19A-O; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 975 AA;

alignment_scores:
Quality: 261.50 Length: 1205
Ratio: 0.498 Gaps: 54
Percent Similarity: 43.568 Percent Identity: 19.253

alignment_block:
US-09-303-518D-651 x AAB01826 ..

Align seg 1/1 to: AAB01826 from: 1 to: 975

```
 892 AAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTCTN 941
     |||:::|||:::                  |||      :::|||::
   1 LysGluTrpLeuLeu.................AspProAspAsnValSe 11

 942 TTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACG 991
     :::|||      |||:::   :::           :::::::::|||
  11 rIleGluAsnProSerThrGluArg...........AsnAspSerAsnG 24

 992 GTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAGCTT 1041
                 |||   |||:::|||:::::::::|||||||||:::
  24 luAspLeuGluTyrThrGlyThrGlyGluAsnIleAsnAsnProLysVal 40

1042 AAAGTACAGACAGTCCGACTGTTTGACGAATCTTTG..............  1077
     :::   |||:::   :::   :::   :::|||:::
  41 AsnAsnGlnSerLysLysThrLeuThrSerSerIleLeuGluAsnIleLe 57

1078 ..................AATGAAACTGATAAAGAACCAGTTTACGCGG 1108
                        |||  |||      :::   :::|||
  57 uLysLysGlySerPheValAsnIleThrAlaThrAspAsnIleTyr.... 72

1109 CAGGGGGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTT 1158
                                  :::|||::::::   |||:::
  73 ............................ValAsnSerSerIleAsnIle 79

1159 TCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACATCAA 1208
     :::              :::|||   ||||||||||||
  80 Gly............AspSerGlyHisLeuIleLeuSer........... 88

1209 CCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTG 1258
          :::||||||
  89 .......GlyGlyGlyArg.............................. 92

1259 AAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGT 1308
                    :::|||:::||||||   |||:::::::::
  93 ................AsnGlyGlyGlyValLysIleAsnLysAsnIle 103

1309 ACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAAT 1358
     |||   |||
 104 ThrSerThr..................................... 106

1359 CGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG......GAAAACCAAG 1402
           |||:::|||   :::::::::||||||      :::   ::::
 107 ....GlyGlySerLeuThrIleAsnSerLysGlyTrpValAspIleHisS 122

1403 GCTCGATCAGCGTGGGC.....................GACGGTACA 1428
     :::::||||||:::|||                      :::|||:::
 122 erAsnIleSerLeuGlyThrGlyPheLeuAsnIleThrSerAsnGlySer 138

1429 GTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAG 1478
     |||   :::::::::          ||||||:::::::::   |||   ::
 139 ValAlaPheGluLysAlaAspLysAspLysAlaArgSerAlaAlaAspAl 155

1479 TGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATA 1528
     ::::|||           ::::::|||   :::::::|||:::::::::|
 155 aGlnIle.........ValAlaGlnGlyIleIleAsnLeuThrGlyGluA 169

1529 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1578
     ||:::                        ||||||   :::::::::
 169 snLys.......................ThrPheArgLeuAsnAsnVal 177

1579 GATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGA 1628
     :::|||||||||                                :::   ::
 178 SerLeuAsnGly............................ValGlyGl 184

1629 AGGGGCGATGATTGNCNATCATAATGCCACAACAACA.....T 1666
     :|||      |||::::::::::   ::::::   |||        :
 184 nGlyLeuSerIleThrSerAsnValGlyAsnGlnThrHisLysPheAspG 201

1667 CCACCGTTACCATTACAGGGAATGAAAGTATTACACAA..... 1704
     ::   ::::::||||||||||||  :::|||:::|||
 201 lyGluIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 217

1705 .....
                                  |||      |||::::::
 218 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 234

1728 ACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAG 1777
      ||||||   :::|||           |||:::   |||
 234 rLeuAsnValGlnLys...........AsnSerSerPheThrPheIleL 247

1778 ATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCA 1827
```

```
                 |||::::::|||                    :::   |||::::::
 247 ysArgThrGluSerAsnArg.............PheGlyProThrThr 258

1828 GAAGACCGCACCCNGCTGCTTTCCGGCGGA.........ACAAATTTAAA 1868
          |||:::          |||||||||          ::::::   ||
 259 ProLeuArgSer.........SerGlyGlyValPhePheAsnGlyThrAs 272

1869 CGGCAACATC.........ACGCAAACAAACGGCAAACTGTTTTTCAGCG 1909
     |||||||:::         :::   ||||||:::::::::::::|||:::
 272 nGlyAsnMetValLeuAsnValGlyThrAsnSerArgValLeuPheAsnL 289

1910 GCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAA 1959
       :::|||:::   ::::::   |||                  ||||||
 289 euLysProAsnGluAsnThrAsnAsn.................SerLys 299

1960 ATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNA 2009
                                                    :::
 300 Pro.........................................LeuPr 302

2010 CCGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGA 2059
           |||:::|||:::          |||   ||||||
 302 oLeuGlnPheAsnAlaAsnIleThrAlaIleGlyGlyGly......... 315

2060 TTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCAC 2109
                    :::|||:::   |||   |||   :::||||||
 316 ................SerValSerPheAspIleHis...AlaAsnHis 325

2110 GCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTAC 2159
:::            |||   :::   :::   :::   :::      ::
 326 Ser............GlyArgGlyAlaGluLeuLysMetAsnThrIleAs 338

2160 ACGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCG 2209
:   ||||||   |||:::|||||       :::::::::   ::: .
 338 nIleSerAspGlyThrSerLeuThr......LeuGlnSerHisValArgL 353

2210 ACGATAAAGTGATT..................................... 2223
  |||::::::::::
 353 ysAspSerAlaPheIleIleSerLysAspLeuThrIleAsnAlaThrGly 369

2223 ..................................................... 2223

 370 SerAsnPheThrLeuGluGlnSerProAspSerPheThrAspLysTyrPr 386

2224 ..........GCTTCATTGACTAAG......ACNGACNTNAGCGGCANTG 2257
               |||   ||||||      |||   ::::::|||   |
 386 oGlyArgAlaIleSerSerThrLysAsnIleThrIleSerGlyGlyAsnV 403

2258 TNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTN 2307
     ||:::|||                         :::::::::   :::
 403 alSerLeuGlyGlyGln..............AsnSerSerSerAspIle 414

2308 AANGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTC...AGCCA 2354
        ||||||::::::   :::::::::|||:::   |||:::   :::||
 415 LysGlyAsnIleThrIleLysSerSerThrAsnValThrLeuLysAlaHi 431

2355 CAACGCCACCCAAAAC.......................GGCAACCTTA 2380
     ||||:::   ::::::                      |||||||||:
 431 sAsnSerProArgAspPheAlaSerArgThrLeuThrLeuGlyAsnLeuA 448

2381 GCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC 2430
     :::::   ||||||         |||   :::   |||   :::::|||
 448 snValGluGlyAsnLeuThrLeuThrGlySerValAlaAspIleLysGly 464

2431 AACNCATCGGNTTCGGGCAATGCTTCATTT..................... 2460
     |||   |||      ::::::|||:::|||
 465 AsnLeuSerIleLeuAsnAspAlaThrPheLysGlyGluThrSerGluAs 481

2460 ..................................................... 2460

 481 nLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaAspIleAsnI 498

2461 ................AATCTAAGCAACAACGCCGCACAAAACGGCAGT 2493
                     |||::::::::::|||   :::::::::|||:::

 498 leLysGlnGlyValValAsnIleGlnGlyAsnIleThrAsnLysGlyGly 514

2494 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAA 2543
     |||:::::::::   ||||||:::   |||   :::   :::   :::||
 515 LeuAsnIleThrThrAsnAlaGlnAsnAsn...GlnLysThrIleIleAs 530

2544 CGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCC 2593
     |||||||::::::
 530 nGlyAsnIleThr..................................... 534

2594 GCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTAAAA 2643
                                 ::::::         |||:::::::|||
 535 .......................AsnGluGlyGlyAspLeuAsnIleLys 543

2644 GACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA.........GGCAA 2684
     |||||||::            :::::::::|||:::         |||||
 544 AspSerAsn.............AsnAsnAlaGluIleGlnIleGlyGlyAs 556

2685 TTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACG 2734
     |:::::      ::::::::::::|||::::::|||            |
 556 nIleSerGlnLysLysGlyAsnLeuThrIleSerSer............A 569

2735 AT..............................GCTGCAGGCGCGCAAACC 2754
     ||                                |||||||::::::
 569 spLysIleAsnIleThrLysLysIleThrIleLysAlaGlyValAspGlu 585

2755 GGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2804
     |||      ::::::::::|||      :::::::::   ::::::
 586 GlyGlySerAspSerSerProAlaSerAsnAlaAsnLeuThrIle..... 600

2805 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2854
                                      :::   :::||||||   ::::
 601 ............................LysThrLysThrLeuGluLeuT 608

2855 ACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTC 2904
     ::|||   ||||||   :::|||      :::      :::|||:::
 608 hrGlyAspLeuAsnIleSerGlyPheAsnLys......AlaGluIleThr 622

2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
     :::   :::::::::   |||   :::::::::::|||:::|||
 623 AlaLysAsnGlyAsnAspLeuThrIleGlyLysAlaSerAspGly..... 637

2955 CACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAAT 3004
                    |||::::::   :::   |||::::::|||::::
 638 ..............AsnAlaAsnAlaLysLysValThrPheAspLysV 649

3005 TGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAAT 3054
     ::               ||||||:::|||      :::::::::   |||
 649 al..............LysAspSerLysIleSerAlaAsnGlyHisAsn 660

3055 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3104
        |||
 661 ValThr.......................................... 662

3105 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3154

 662 ..................................................... 662

3155 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3204
       |||:::::::|||:::      :::::::::   ::::::|||:::
 663 ..LeuAsnSerLysValGluThrSerAsnSerAspSerSerAlaAspAsp 678

3205 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGC 3254
     :::|||   ::::::              |||:::|||   :::|||:::::
 679 SerAsnAspAsnAsnThrGlyLeuThrIleSerAla...LysAspValTh 694

3255 CGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAA 3304
     :
 694 r..................................................... 694

3305 ATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGAT 3354
                                                 |||:::   |||
 695 ..................................ValAsnAsnAsp 698
```

```
 3355 AAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTAC 3404
          |||      :::   :::   :::|||   |||         ||
  699 ValThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnValTh 715

 3405 CACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGC 3454
      ||||
  715 rThr..................................................... 716

 3455 AGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGT 3504

  716 ..................................................... 716

 3505 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTT 3554
                          :::   |||:::|||:::||| |||
  717 .....................LysGluSerThrThrIleAsn.AlaAla. 725

 3555 CGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACG 3604
                         |||:::   :::      ||||||   |||
  726 ....................ThrGlySerValGluValThrAlaLysThr 735

 3605 CNGTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGCAAGAT 3654
                       |||||||||                    ||
  736 Gly.........AspIleSerGlyThr.....................Il 742

 3655 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAA 3704
      ||||:::   |||:::|||     |||:::::::   |||:::      :
  742 eSerGlyAsnThrValAsn...ValThrAlaThrAspSerLeuThrThrG 758

 3705 AAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCG 3754
      ::               ||||||:::|||         :::::::|||
  758 ln..............AlaSerSerSerIleThrSerSerAsnGlyGln 769

 3755 AAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3804
         |||   ::::::      ::::::         ::::::   :::||
  770 ThrThrLeuThrAlaLysAsnGlySerIleAlaGlySerIleAspAlaAl 786

 3805 GCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCAC 3854
      |      :::   ||||||::::::::   :::|||   |||:::::::
  786 aAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerA 803

 3855 GGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAA 3904
      ::   ::::::   :::::::::   :::   :::|||:::::||||||
  803 snIleLysAlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLys 819

 3905 TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 3954
            :::::::::                        |||   :::  ||
  820 LeuAspGlyThrAla..............SerGlyAspArgThrValVa 831

 3955 TTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGT 4004
      |:::|||   :::||||||:::     |||:::      |||   ||||
  831 lAsnAlaThrAsnAlaSerGlySerGlySerValThrAlaAlaThrSerS 848

 4005 CCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTC 4054
      ||:::   :::|||||||:::      :::::::|||:::
  848 erAsnValAsnIleThrGlyAspLeuSerThrIleAsnGly......... 861

 4055 TTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAA 4104
                                       :::||||||   :::||
  862 ............................LeuAsnIleIleSerLysAs 868

 4105 CCGGCGCAACACA 4117
      |   :::||||||
  868 nGlyLysAsnThr 872
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827

seq_documentation_block:

```
ID   AAB01827 standard; Protein; 969 AA.
XX
AC   AAB01827;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc mature HMW2A protein, SEQ ID NO:32.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52178.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 19A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   969 AA;
```

```
alignment_scores:
           Quality:  261.00              Length:   1169
             Ratio:  0.505                 Gaps:     53
 Percent Similarity:  44.226   Percent Identity: 19.247

alignment_block:
 US-09-303-518D-651 x AAB01827   ..

 Align seg 1/1  to: AAB01827  from: 1  to: 969
```


1015 AACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTT 1064
3 AsnValSerIleGluAsnProSerThrGlu 12
1065 TGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCGGCAGGGG 1114
13 ArgAsnAspSerAsnGluAspLeuGluTyrThrGlyThrG 26
1115 GTGTTAATCAGTACCGTCCAAGGTTAAACAACGGT 1149
26 lyGluAsnIleAsnAsnProLysValAsnAsnGlnSerLysLysThrLeu 42
1150 GAAAACCTT TCTTTTATCGATTA 1172
43 ThrSerSerIleLeuGluAsnIleLeuLysLysGlySerPheValAsnIl 59
1173 C...GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGC...G 1216
59 eThrAlaThrAspAsnIleTyrValAsnSerSerIleAsnIleGlyAspS 76
1217 CGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAAAACAAC 1266
76 erGlyHisLeuIleLeuSerGlyGlyGlyArg 86
1267 GAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTAC 1316
87 AsnGlyGlyGlyValLysIleAsnLysAsnIleThrSerTh 100
1317 TTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAG 1366
100 r GlyG 102
1367 GCACGCTGCACGTTCAAGCCAAAGGG GAAAACCAAGGCTCGATC 1410
102 lySerLeuThrIleAsnSerLysGlyTrpValAspIleHisSerAsnIle 118
1411 AGCGTGGGC GACGGTACAGTCATTTT 1436
119 SerLeuGlyThrGlyPheLeuAsnIleThrSerAsnGlySerValAlaPh 135
1437 GGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCG 1486
135 eGluLysAlaAspLysAspLysAlaArgSerAlaAlaAspAlaGlnIle. 151
1487 GCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTC 1536
152 ValAlaGlnGlyIleIleAsnLeuThrGlyGluAsnLys... 164
1537 AACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAA 1586
165 ThrPheArgLeuAsnAsnValSerLeuAs 174
1587 CGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGA 1636
174 nGly ValGlyGlnGlyLeuS 181
1637 TGATTGNCNATCATAATGCCACAACAACA TCCACCGTT 1674
181 erIleThrSerAsnValGlyAsnGlnThrHisLysPheAspGlyGluIle 197
1675 ACCATTACAGGGAATGAAAGTATTACACAA 1704
198 AsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAlaThrThrAl 214
1705 CCGAGTGGTAAGAATATCAATAGACTTAATT 1735
214 aTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerThrLeuAsnV 231
1736 ACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACC 1785
231 alGlnLys AsnSerSerPheThrPheIleLysArgThr 243
1786 AAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCG 1835
244 GluSerAsnArg PheGlyProThrThrProLeuAr 255
1836 CACCCNGCTGCTTTCCGGCGGA ACAAATTTAAACGGCAACA 1876
255 gSer SerGlyGlyValPhePheAsnGlyThrAsnGlyAsnM 269
1877 TC ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCG 1917
269 etValLeuAsnValGlyThrAsnSerArgValLeuPheAsnLeuLysPro 285
1918 ACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGG 1967
286 AsnGluAsnThrAsnAsn SerLysPro 294
1968 TATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGCACGT 2017
295 LeuProLeuGlnP 299
2018 TTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGC 2067
299 heAsnAlaAsnIleThrAlaIleGlyGlyGly 309
2068 AATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGC 2117
310 SerValSerPheAspIleHis...AlaAsnHisSer 320
2118 AGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGG 2167
321 GlyArgGlyAlaGluLeuLysMetAsnThrIleAsnIleSerA 335
2168 ACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAA 2217
335 spGlyThrSerLeuThr LeuGlnSerHisValArgLysAspSer 349
2218 GTGATT 2223
350 AlaPheIleIleSerLysAspLeuThrIleAsnAlaThrGlySerAsnPh 366
2223 2223
366 eThrLeuGluGlnSerProAspSerPheThrAspLysTyrProGlyArgA 383
2224 ..GCTTCATTGACTAAG ACNGACNTNAGCGGCANTGTNAGNCTN 2265
383 laIleSerSerThrLysAsnIleThrIleSerGlyGlyAsnValSerLeu 399
2266 NCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAA 2315
400 GlyGlyGln AsnSerSerSerAspIleLysGlyAs 411
2316 TCTTAGTGCAAATGGCGATACACGTTATACAGTC...AGCCACAACGCCA 2362
411 nIleThrIleLysSerSerThrAsnValThrLeuLysAlaHisAsnSerP 428
2363 CCCAAAAC GGCAACCTTAGCCTCGTG 2388
428 roArgAspPheAlaSerArgThrLeuThrLeuGlyAsnLeuAsnValGlu 444
2389 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACNCATC 2438
445 GlyAsnLeuThrLeuThrGlySerValAlaAspIleLysGlyAsnLeuSe 461
2439 GGNTTCGGGCAATGCTTCATTT 2460
461 rIleLeuAsnAspAlaThrPheLysGlyGluThrSerGluAsnLeuAsnI 478
2460 2460
478 leThrGlyAsnPheThrAsnAsnGlyThrAlaAspIleAsnIleLysGln 494
2461 AATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCT 2501

```
 495 GlyValValAsnIleGlnGlyAsnIleThrAsnLysGlyGlyLeuAsnIl 511

2502 TTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATG 2551
     ::::   ||||||:::   |||   :::   :::   :::|||||||||:
 511 eThrThrAsnAlaGlnAsnAsn...GlnLysThrIleIleAsnGlyAsnI 527

2552 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACC 2601
     :::::
 527 leThr..................................................... 528

2602 GGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGA 2651
                  ::::::         |||::::::|||||||||::
 529 ..............AsnGluGlyGlyAspLeuAsnIleLysAspSerAs 540

2652 ATGGACGCTGCCGTCAGGCACGGAATTA.........GGCAATTTAAACC 2692
     :           :::::::::|||:::          ||||||::::::
 540 n............AsnAsnAlaGluIleGlnIleGlyGlyAsnIleSerG 553

2693 TTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGAT...... 2736
          :::::::::::::|||::::::|||            |||
 553 lnLysLysGlyAsnLeuThrIleSerSer............AspLysIle 565

2737 .......................GCTGCAGGCGCGCAAACCGGCAGNGT 2762
                             ||||||::::::   |||
 566 AsnIleThrLysLysIleThrIleLysAlaGlyValAspGluGlyGlySe 582

2763 GTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2812
      :::::::::|||      :::::::::   ::::::
 582 rAspSerSerProAlaSerAsnAlaAsnLeuThrIle.............. 594

2813 CGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAA 2862
                       :::   :::|||||||   :::::::|||
 595 ....................LysThrLysThrLeuGluLeuThrGlyAsp 604

2863 TTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCG 2912
     ||||||   :::|||      :::      :::|||:::   :::   ::
 605 LeuAsnIleSerGlyPheAsnLys......AlaGluIleThrAlaLysAs 619

2913 AAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGG 2962
     :::::::   |||   :::::::::::::|||:::|||
 619 nGlyAsnAspLeuThrIleGlyLysAlaSerAspGly.............. 631

2963 CGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTA 3012
               |||:::::::   :::   |||:::::::|||::::::
 632 ........AsnAlaAsnAlaLysLysValThrPheAspLysVal...... 643

3013 GTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCT 3062
               ||||||:::|||      :::::::::   |||   |||
 644 .........LysAspSerLysIleSerAlaAsnGlyHisAsnValThr.. 656

3063 GCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCA 3112

 656 ..................................................... 656

3113 AAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCC 3162
                                                  |||:::
 657 ..............................................LeuAsn 658

3163 GACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGC 3212
     :::|||:::     :::::::::   ::::::|||:::   :::|||
 659 SerLysValGluThrSerAsnSerAspSerSerAlaAspAspSerAsnAs 675

3213 GCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGA 3262
      ::::::           |||:::|||   :::|||::::::
 675 pAsnAsnThrGlyLeuThrIleSerAla...LysAspValThr....... 688

3263 CAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGC 3312

 688 ..................................................... 688

3313 ATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAG 3362
                                 |||:::   |||      ||
 689 ............................ValAsnAsnAspValThrSe 695
```

```
3363 CGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGCCT 3412
       |      :::   :::   :::|||   |||         ||||||
 695 rHisLysThrIleAsnIleSerAlaThrThrGlyAsnValThrThr.... 710

3413 TCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAA 3462

 710 ..................................................... 710

3463 CCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAA 3512

 710 ..................................................... 710

3513 TAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTAC 3562
                  :::   |||:::|||:::||| |||
 711 ............LysGluSerThrThrIleAsn.AlaAla......... 719

3563 AGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGG 3612
                  |||:::   :::       ||||||   |||
 720 ............ThrGlySerValGluValThrAlaLysThrGly..... 730

3613 ACAAGCNGCATCCGGNACACCAAACACTACCGTTCGCAAGATTTCCGCGC 3662
               |||||||||                        ||||||:::
 731 ....AspIleSerGlyThr.....................IleSerGlyA 739

3663 CTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCG 3712
       |||:::|||      |||::::::   |||:::      :::
 739 snThrValAsn...ValThrAlaThrAspSerLeuThrThrGln...... 752

3713 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANC 3762
               ||||||:::|||         ::::::|||      |||
 753 .........AlaSerSerSerIleThrSerSerAsnGlyGlnThrThrLe 766

3763 TTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTT 3812
      ::::::      ::::::         ::::::   :::|||      :
 766 uThrAlaLysAsnGlySerIleAlaGlySerIleAspAlaAlaAsnValT 783

3813 CGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGG 3862
     ::   |||||:::::::::   :::|||   |||:::::::::   :::
 783 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAsnIleLys 799

3863 GTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGC 3912
     :::   :::::::::   :::   :::|||::::::|||||      ::
 800 AlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLysLeuAspGl 816

3913 CGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGG 3962
     :::::::                  |||   :::   |||:::|||
 816 yThrAla..............SerGlyAspArgThrValValAsnAlaT 828

3963 ATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAG 4012
       :::||||||:::      |||:::      |||   ||||||:::
 828 hrAsnAlaSerGlySerGlySerValThrAlaAlaThrSerSerAsnVal 844

4013 CGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTC 4062
     :::|||||||:::      ::::::|||:::
 845 AsnIleThrGlyAspLeuSerThrIleAsnGly.................. 855

4063 AACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4112
                                 :::|||||   :::|||   :::|
 856 .....................LeuAsnIleIleSerLysAsnGlyLysA 865

4113 ACACA 4117
     |||||
 865 snThr 866
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120

seq_documentation_block:
ID AAU37120 standard; Protein; 2344 AA.
XX
AC AAU37120;
XX
DT 14-FEB-2002 (first entry)

```
XX
DE   Staphylococcus aureus cellular proliferation protein #1290.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54979.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12713; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2344 AA;


alignment_scores:
            Quality:  260.00              Length:   1577
              Ratio:  0.344                Gaps:     59
 Percent Similarity:  47.876   Percent Identity: 18.199

alignment_block:
 US-09-303-518D-651 x AAU37120   ..

 Align seg 1/1  to: AAU37120  from: 1  to: 2344

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
         |||   :::   :::::::|||:::            |||   |||:::||
      20 ArgLeuTyrLysSerGlyLysAsnTrpValLysSerGlyIleLysGluIl 36

     213 TGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAG 262
         ||||::::::::      |||
      36 eGluMetPheLysIleMetGly.........................L 44

     263 CCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC..... 300
           |||:::|||:::   |||:::||||||:::::::::
      44 euProPheIleSerHisSerMetValSerGlnAspAsnGlnSerIleSer 60

     301 .........................GTGGCGGCATTGGTGGGCGATCA 323
                                 ::::::|||::::::|||
      61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyGlyAl 77

     324 ATATATTGTGAGCGTGGCACATAAC..................... 348
          :::   |||::::::   |||:::
      77 aPheThrValAsnMetLeuHisAspGlnGlnAlaPheAlaAlaSerAspA 94

     349 .....................................GGCGGCTAT 357
                                             |||:::
      94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110

     358 AACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTT 407
         |||::::::   :::   |||:::   ::::::   |||:::
     111 AsnSerThrThrIleAspAlaSerThrSerThrAlaAspSer........ 124

     408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACC 457
                       |||   :::|||:::   :::   ::::::|||
     125 .......ThrSerValThrLysAsnSerSerSerValGlnThrSer.... 137

     458 CTTACAACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCACAGAT 507
               |||:::|||:::         :::      :::   |||||||:::
     138 .....AsnSerAspThrValSerSerGluLysSerGluAsnValThrSer 152

     508 ....................................GCAGAACCTGTCGA 521
                                            ::::::      ||
     153 ThrThrAsnSerThrSerAsnGlnGlnGluLysLeuThrSerThrSerGl 169

     522 AATGACGAGTGACATGAGGGGGAATACCTATTCCGATAAAGAAAAATATC 571
         |   ||||||:::   :::   ::::::   ||||||    |||:::::::
     169 uSerThrSerSerLysAsnThrThrSerSerSerAsp.ThrLysSerVal 185

     572 CCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGAT 621
            |||:::|||::::::   |||      :::   ::::::   :::
     186 ThrSerThrSerSerThrGluGlnProIleAsnThrSerThrAsnGlnSe 202

     622 GACA.........AACACGGCGATTTATCCTACTCCGGCGCATGGTTAAT 662
          |||          ||||||:::          ||||||:::   :::   |
     202 rThrAlaSerAsnAsnThrSerGlnSerThrThrProThrSerAlaAsnL 219

     663 TG............................GCGGCAATACACATATGC 682
         ||                           |||   :::   :::
     219 euAsnLysThrSerThrThrSerThrSerThrAlaProValLysLeu... 234

     683 AGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCAT 732
                   :::   ::::::   |||:::   ::::::::::::::
     235 .........ArgThrPheSerArgLeuAlaMetSerThrPheAlaSerAl 248

     733 GCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGG.............. 769
               ||||||   ||||||          :::
     248 aAlaThrThrThrAlaLeuThrAlaAsnThrIleThrValAsnLysAspA 265

     770 ........................................CAGGCG 775
                                                 |||:::
     265 snLeuLysGlnTyrMetThrThrSerGlyAsnAlaThrTyrAspGlnSer 281

     776 ACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTG 825
         |||  |||          ::       |||   |||:::::::|||
     282 Thr..GlyValValThrLeuThrGlnAspThrThrSerGlnLysGlyAla 297

     826 CTC.........................................AACGG 833
         :::                                         :::||
     298 IleThrLeuGlyThrArgIleAspSerAsnLysSerPheHisPheSerGl 314

     834 AGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGC 883
         |   ::::::   |||      |||:::|||:::   ||||||
```

```
 314 yLysValAsnLeuGlyAsnLysTyrGluGlyHisGlyAsnGly..... 328

 884 TGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACAT 933
                                        ||||||
 329 ..................................GlyAspGlyIle 332

 934 ACCGTCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAA 983
       :::|||:::|||         |||:::   ::::::::::::
 333 GlyPheAlaPheSerProGlyValLeuGlyGluThrGlyLeuAsnGlyAl 349

 984 C...................AACAACGGTACGGGTACGGTAACAGAAA 1012
                          :::|||:::   |||         :::|
 349 aAlaValGlyIleGlyGlyLeuSerAsnAlaPheGlyPheLysLeuAspT 366

1013 CCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTG 1062
     ||   :::::::::|||:::|||
 366 hrTyrHisAsnThrSerThrPro........................... 373

1063 TTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCGGCAGG 1112
       :::|||      ::::::      :::|||      :::||||||
 374 ......AsnSerSerAlaLysAlaLysAlaAspProSerAsnValAlaGl 388

1113 GGGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTT 1162
     ||||                                |||         :::|
 388 yGly.............................GlyAlaPheGlyAlaP 395

1163 TTATC.........GATTACGGCAACGGCAAACTCATCTTATCAAACAAC 1203
     ||:::         :::||||||   ::::::          |||::::::
 395 heValThrThrAspSerTyrGlyValAlaSerThrTyrThrSerSerSer 411

1204 ATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTC 1253
       ::::::|||:::   |||                    :::|||::
 412 ThrAlaAspAsnAlaAlaLysLeu..................AsnValGl 422

1254 GCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAG 1303
     :|||   ||||||   |||:::|||         ::::::   :::   |
 422 nProThrAsnAsn...ThrPheGlnAspPheAspIleAsnTyrAsnGlyA 438

1304 ACAGTACCGTT.......................ACTTGGAAAGTAAAC 1329
     ||:::   |||                      ||||||      |||
 438 spThrLysValMetThrValLysTyrAlaGlyGlnThrTrpThrArgAsn 454

1330 GGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGT 1379
           :::|||   :::::::|||   |||      ::::::   ::
 455 ......IleSerAspTrpIleAlaLysSerGlyThrThrAsnPheSerLe 469

1380 TCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC.........GTGGGCG 1420
     ::::      :::::::::   |||:::   :::         |||
 469 uSerMetThrAlaSerThrGlyGlyAlaThrAsnLeuGlnGlnValGlnP 486

1421 ACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAA 1470
       ||||||          :::::::|||      :::   :::
 486 heGlyThrPheGluTyrThrGluSerAlaValThrGlnValArg...... 500

1471 GCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGG.................. 1503
                :::   :::::::::|||:::
 501 .........TyrValAspValThrThrGlyLysAspIleIleProProLy 514

1504 ..........GGTACGGTGCAA......CTGAATGCCGATAATCAG.... 1533
               |||:::|||:::      ::::::   |||||||||
 514 sThrTyrSerGlyAsnValAspGlnValValThrIleAspAsnGlnGlnS 531

1534 ....................TTCAACCCCGACAAACTCTATTTCGGCTTT 1563
                         :::|||      ::::::      ::::::
 531 erAlaLeuThrAlaLysGlyTyrAsnTyrThrSerValAspSerSerTyr 547

1564 CGCGGCGGACGTTTGGATTTAAAC.........................GG 1589
       :::         |||   |||                          ||
 548 AlaSerThrTyrAsnAspThrAsnLysThrValLysMetThrAsnAlaGl 564

1590 GCATTCGCTTTCGTTCCACCGTATT.........................1614
     |:::|||:::::::::::
 564 yGlnSerValThrTyrTyrPheThrAspValLysAlaProThrValThrV 581

1615 ........CAAAATACCGATGAAGGG..................GCGATG 1638
               |||:::   :::   |||                 ::::::
 581 alGlyAsnGlnThrIleGluValGlyLysThrMetAsnProIleValLeu 597

1639 ATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGAA 1688
       ::::::   |||:::|||   |||   |||   |||:::||||||
 598 ThrThrThrAspAsnGlyThrGlyThrValThrAsnThrValThrGly.. 613

1689 TGAAAGTATTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTACA 1738
                  |||||||||                 |||:::|||
 614 .............LeuProSerGly..............LeuSerTyr. 620

1739 GCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAA............ 1776
       :::   |||   |||:::   :::|||
 621 .....AspSerAlaThrAsnSerIleIleGlyThrProThrLysIleGly 635

1777 ....................GATACGACCAAAACGAACGGGCGG..... 1800
                         :::|||   ::::::|||::::::
 636 GlnSerThrValThrValValSerThrAspGlnAlaAsnAsnLysSerTh 652

1801 .............CTCAACCTTGTT.................TACCAGC 1819
                  :::|||:::|||                  |
 652 rThrThrPheThrIleAsnValValAspThrThrAlaProThrValThrP 669

1820 CCGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAAC 1869
     ||   ::::::      :::   :::::::|||      :::   :::|||
 669 roIleGlyAspLysSerSerGluValPheSerProIleSerProIleAsn 685

1870 GGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1919
               ||||||   |||                  |||||||:::   ::
 686 ...IleAlaThrGlnAspAsn..............SerGlyAsnAlaVa 696

1920 ACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTA 1969
     :   ::::::                                :::   |||:
 696 lThrAsnThr............................ValThrGlyL 703

1970 TCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGCACGTTT 2019
     ::|||:::|||                                ||||||
 703 euProSerGly............................LeuThrPhe 709

2020 AAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAA 2069
       :::   |||      |||:::|||      :::   |||:::   ::
 710 AspSerThrAsnAsnThrIleSerGlyThrProThrAsnIleGlyThrSe 726

2070 TGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAG 2119
     ::::      :::   :::            :::|||   :::   ::::
 726 rThrIleThrIleValSerThrAspAlaSerGlyAsnLysThrThrThrT 743

2120 TTTTTGGT.........GTCGCACCGCATCAAAGCCATACAATCTGTACA 2160
     ::|||          :::      :::|||   ::::::   |||
 743 hrPheLysTyrGluValThrArgAsnSerMetSerAspSerValSerThr 759

2161 CGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGA 2210
                ::::|||   |||:::      :::   :::::::::
 760 ............SerGlySerThrGln...GlnSerGlnSerValSerTh 771

2211 CGATAAAGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNA 2260
     :::|||:::   ::::::   :::::::|||:::::|||||||:::::
 771 rSerLysAlaAspSerGlnSerAlaSerThrSerThrSerGlySerIleM 788

2261 GNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAAN 2310
          :::::::::::::::::::::   :::   :::::::::|||:::
 788 etThrSerThrSerAlaSerThrSerLysSerThrSerValSerLeuSer 804

2311 GGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGC 2360
       ::::::|||||||:::   :::         :::   |||::::::::
 805 AspSerValSerAlaSerLysSerLeuSerThrSerGluSerAsnSerVa 821

2361 CACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2410
     ::::::::::   :::   |||||||||   |||:::|||::::::
 821 lSerSerSerThrSerThrSerLeuVal...AsnSerGlnSerVal.... 835
```

```
2411 ATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTT 2460
836 ..SerSerSerMetSerGlySerValSerLysSerThrSerLeuSerAsp 851

2461 AATCTAAGCAACAACGCC.....GCACAAAACGGCAGTCTGAC 2498
852 PheIleSerAsnSerSerSerThrGluLysSerGluSerValSerThrSe 868

2499 GCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCC..... 2535
868 rThrSerAspSerLeuArgThrSerThrSerLeuSerAspSerValSerM 885

2536 ..GCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTT 2583
885 etSerThrSerGlySerLeuSerLysSerGlnSerLeuSerThrSerThr 901

2584 GAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATT 2633
902 SerAspSerAlaSerThrSerGln...SerValSerAspSerThrSerAs 917

2634 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCA 2683
917 nSerIleSerThrSerGluSerLeuSerGluSerGlySerThrSerGluS 934

2684 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2733
934 erIleSerIleSerAsnSerIleSerAsnSerValSerAlaSerThrSer 950

2734 GATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCG 2783
951 LysLeuGluSerGlnSerThrSerIleSerLeuSerThrSerAspSerLy 967

2784 TTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCC 2833
967 sSerMetSerThrSerGluSerLeuSerAspSerThrSerThrSerAsp. 983

2834 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTC 2883
984 .....SerValSerGlySerLeuSerValAlaGlySerGln 995

2884 CGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGC 2933
996 SerValSerThrSerThrSerAspSerMetSerThrSerGluMetIleSe 1012

2934 GGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCAACAATACCGGCAACG 2983
1012 rAspSerMetSerThrSerGlySerLeuAlaAlaSerAspSerLysSerM 1029

2984 AACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAA 3033
1029 etSerValSerSerSerMetSerThrSerGlnSerGlySerThrSerGlu 1045

3034 CCGCTGTCCGAAAACCTT.....AATTTCAC 3059
1046 SerLeuSerAspSerIleSerThrSerAspSerAspSerLysSerLeuSe 1062

3060 CCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCC 3109
1062 rLeu.SerThrSerGlnSer.....GlySerThrSerThrSer 1074

3110 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3158
1075 ThrSerThrSerSerSerValArgMetSerGluSerGlnSerThrSerGl 1091

3159 .TCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACA 3208
1091 ySerMetSerThrSerGlnSerAspSerThrSerIleSerThrSerPheS 1108

3209 ACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAA 3258
1108 erAspSerThrSerAspSerLysSerAlaSerThrAlaSerSerGluSer 1124

3259 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGT 3308

1125 IleSerGlnSerValSerThrSerThrSerGlySerValSerThrSerTh 1141

3309 CGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAG 3358
1141 rSerLeuSerThrSerAsnSerGluArgThrSerThrSerMetSerAspS 1158

3359 ACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACC... 3405
1158 erThrSerLeuSerThrSerGluSerAspSerThrSerAspSerThrSer 1174

3405 ..... 3405
1175 ThrSerAspSerIleSerGluAlaIleSerGlySerGluSerThrSerIl 1191

3406 .....ACCGCCTTCCCCCGCGCCCGCNGCGCCC 3433
1191 eSerLeuSerGluSerAsnSerThrSerAspSerGluSerLysSerAlaS 1208

3434 GCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCG 3483
1208 erAlaPheLeuSerGluSerLeuSerGluSerThrSerGluSerThrSer 1224

3484 CAGCGCGACCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTC 3533
1225 GluSerLeuSerGlySerThrSerAspSerThrSerLeuSerAspSerAs 1241

3534 CGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGT 3583
1241 nSer.GluSerGlySerThrSerThrSerLeuSerAsnSerThrSerGly 1257

3584 TTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGNACA.. 3631
1258 SerAlaSerIleSerThrSerThrSerGlySerAlaSerThrSerThrVa 1274

3631 ..... 3631
1274 lLysSerGluSerValSerThrSerLeuSerThrSerThrSerThrSerL 1291

3632 .....CCAAACACTACCGTT 3646
1291 euSerAspSerThrSerLeuSerThrSerLeuSerAspSerThrSerGly 1307

3647 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3696
1308 SerLysSerAsnSerLeuSerAlaSerMetSerThrSerAspSerIleSe 1324

3697 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAA 3746
1324 rThrArgLysSerGluSerLeuSerAlaSerThrSerLeuSerGlySerT 1341

3747 CCGGA...CCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGC 3793
1341 hrSerGluSerGluSerGlySerThrSerSerSerGluSerLysSerAsp 1357

3794 TTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3843
1358 SerThrSerMetSerLeuSerMetSerGlnSerThrSerGlySerThrSe 1374

3844 GGCATCAGCACGGGCGCGGGTT.....T 3866
1374 rValSerThrSerGluSerLeuSerAspSerThrSerThrSerLeuSerL 1391

3867 TAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCG 3916
1391 euSerAlaSerMetAsnGlnSerGlyValAspSerAsnSerAlaSerGln 1407

3917 TGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTC 3966
1408 SerAlaSerThrSerThrSerThrSerGluSerAspSerGlnSe 1424

3967 GGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGA 4016
```

```
     1424 rThrSerSerTyrThrSerGlnSer.........ThrSerGlnSerGluS 1438

     4017 TTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACC 4066
          |||:::|||:::|||||||:::|||                  ||||||
     1438 erThrSerThrSerThrSerLeuSerAsp..............SerThr 1449

     4067 GNTACCGNGCGGGCATTAAGGCAG 4090
          :::   :::::::::      |||
     1450 SerIleSerLysSerThrSerGln 1457

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724

seq_documentation_block:
ID   AAR41724 standard; Protein; 1477 AA.
XX
AC   AAR41724;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 2 (HMW2).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49507.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties .
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1477 AA;


alignment_scores:
           Quality:  259.50            Length:   1395
             Ratio:  0.425               Gaps:     63
 Percent Similarity:  43.728   Percent Identity: 19.068

alignment_block:
 US-09-303-518D-651 x AAR41724   ..

 Align seg 1/1  to: AAR41724  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACAA 227
         |||:::||||||:::::::::         |||            :::||
     282 AsnGlnGlyLysLeuSerAla.........AspSerValSerLysAspLy 295

     228 AAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATT 277
         |:::|||:::::|||   ::::::                    |||
     295 sSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleGlyG 312

     278 TTTCTGTGGTGTCGCGTAAC..............GGCGTGGCGGCATTG 312
               :::   :::::|||                |||        :::
     312 lyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMetIle 328

     313 GTGGGCGATCAATATATTGTGAGCGTG.....................GC 341
         :::||||||:::      :::::::::                     ::
     329 ThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSerGl 345

     342 ACATAACGGCGGC......TATAACAACGTTGATTTTGGTGCGGAAGGAA 385
         :   :::||||||      |||   :::   |||      :::||||||
     345 yLysGluGlyGlyGluThrTyrLeuGlyGlyAspGluArgGlyGluGlyL 362

     386 GNAATCCCGATCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
           |||                  :::   |||:::::|||:::
     362 ysAsn.................GlyIleGlnLeuAlaLysLys...... 370

     436 TATAAGCCTGACAATTCACACCCTTACAACGGCGATTANCATATGCCGCG 485

     370 ..................................................... 370

     486 TTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACA 535
                                                    ||||||
     371 ..........................................ThrSerLeuG 374

     536 TGAGGGGGAATACC......TATTCCGATAAAGAAAAATATCCCGAGCGT 579
           :::|||:::|||         |||   |||||||||
     374 luLysGlySerThrIleAsnValSerGlyLysGluLys............ 386

     580 GTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAACA 629
                 |||:::            |||
     387 .........GlyGlyArgAlaIleValTrp..................... 393

     630 CGGCGATTTATCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATA 679
          ||||||::::::                ||||||   ||||||   :::
     394 .GlyAspIleAla..............LeuIleAspGlyAsnIleAsnA 405

     680 TGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGC 729
           ||||||                            ||||||||||:::
     405 laGlnGly.........................SerGlyAspIleAla 412

     730 CATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAG 779
           :::                          ::::||   :::      ||
     413 LysThr.....................GlyGlyPheValGluThrSe 421

     780 CGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCA 829
         ||||      :::|||||||   |||                      |
     421 rGlyHisAspLeuPheIleLysAsp.....................A 430

     830 ACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 879
         ||:::::::::::::::
     430 snAlaIleValAspAla............................... 435

     880 CAGCTGATACGCAAAGATTGGTTCTACGAT..................... 909
                 |||:::|||:::   |||
     436 ............LysGluTrpLeuLeuAspPheAspAsnValSerIleAs 448

     910 .......GACATTTACAGAGGCGATACACATACCGTCTNTTTTGAACCGC 952
               |||      |||:::::|||                    |||
     448 nAlaGluAspProLeuArgAsnAsnThrGlyIleAsnAspGluPheProT 465

     953 GCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGTACG 1002
           ::::::|||:::   |||      ::::::|||:::         |||
     465 hrGlyThrGlyGluAlaSerAspProLysLysAsnSerGluLeuLysThr 481

    1003 GTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGAC 1052
           |||   ||||||      :::||||||         ::::::   ||
     482 ...ThrLeuThrAsnThrThrIleSerAsnTyrLeuLysAsnAlaTrpTh 497

    1053 AGTCCGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT 1102
         |:::               :::::::::   :::   |||      |||
     497 rMet..............AsnIleThrAlaSerArgLysLeuThrValA 509
```



```
 952 .........................................ThrThr 953

2773 CCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTC 2822
          :::|||::::::||||||:::::::::
 954 HisAlaLysArgAsnGlnArgSerIleIleGly..................... 964

2823 GGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTC 2872
                    :::   ::::::   :::|||:::||||||
 965 .............GlyAspIleIleAsnLysLysGlySerLeuAsnIleT 977

2873 AAGGAACATTCCGCTTTATGTCG...GAACTCTTCGGCTACCGAAGCGAC 2919
       :::   :::      :::   ::::::   |||      |||:::
 977 hrAspSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGln 993

2920 AAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCG..... 2964
     |||                    |||||||:::   |||::::::
 994 Lys.....................GluGlyAsnLeuThrIleSerSerAs 1003

2965 ....GTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGG 3010
         :::|||   |||                        |||:::|||:
1003 pLysIleAsnIleThr.....................LysGlnIleThrI 1013

3011 TAGTGGAAGGGAAAGACAACAAACCGCTGTCC............GAAAAC 3048
     ::   :::|||   |||::::::      |||            :::|||
1013 leLysLysGlyIleAspGlyGluAspSerSerSerAspAlaThrSerAsn 1029

3049 CTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTA 3098
        |||:::|||::::::
1030 AlaAsnLeuThrIleLys............................... 1035

3099 CCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAG 3148
                                              :::||||
1036 .........................................ThrLysG 1038

3149 AACAAGAGCTTTCCGACAAACTC..............GGCAAGGCAGAA 3183
     ||   :::|||::::::   |||              :::|||||||||
1038 luLeuLysLeuThrGluAspLeuSerIleSerGlyPheAsnLysAlaGlu 1054

3184 GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGAT 3233
                  :::   ||||||
1055 Ile.........ThrAlaLysAsp........................... 1059

3234 TGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGG 3283
           |||||||||
1060 .......GlyArgAspLeu............................... 1063

3284 CCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAG 3333
        :::   |||::::::|||   |||            ||||||   :::
1064 .....ThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLys 1078

3334 AAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGA 3383
                  :::      |||||||||   ::::::   :::   ::
1079 ThrValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHi 1095

3384 AGCGGAAACC.........CGGCCGGNTACCACCGCCTTCCCCCGCGCCC 3424
     :      |||          :::      |||::::::          :::|
1095 sAsnValThrLeuAsnSerLysValLysThrSerSerSerAsnGlyGlyA 1112

3425 GCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAA 3474
     ||   ::::::   |||                                :::
1112 rgGluSerAsnSerAspAsnAspThrGlyLeuThrIleThrAlaLysAsn 1128

3475 CCCCAACCGCAGCGCGACCTGATNAGC...CGTTATGCCAATAGCGGTTT 3521
        :::   ::::::|||::::::|||   :::   :::|||
1129 ValGluValAsnLysAspIleThrSerLeuLysThrValAsnIleThrAl 1145

3522 GAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAAT 3571
      ||||||      :::||| :::::::::   :::         ||||||
1145 aSerGluLysValThrThr.ThrAlaGlySerThrIleAsnAlaThrAsn 1161

3572 TGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGC 3621
           |||   :::         ::::::         |||      ::
1162 GlyLysAlaSerIleThrThrLysThrGlyAspIleSerGlyThrIleSe 1178

3622 ATCCGGNACACCAAACACTACCGTTCGC...AAGATTTCCGCGCCTACCG 3668
     ::::::::|||   ::::::::::::         :::::::::   ::::
1178 rGlyAsnThrValSerValSerAlaThrValAspLeuThrThrLysSerG 1195

3669 CCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCG 3718
     :::::|||      :::   ::::::   ::::::   ||||||||||:::
1195 lySerLysIleGluAlaLysSerGlyGluAlaAsnValThrSerAlaThr 1211

3719 GGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGAC 3768
     |||:::   ::::::         |||   :::   :::|||   :::::
1212 GlyThrIleGlyGly.........ThrIleSerGlyAsnThrValAsnVa 1225

3769 GACGGCATCGGCA.................................A 3782
     :||||||:::|||                                :
1225 lThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnA 1242

3783 CTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT......ACGGCA 3826
     ::   :::|||      |||          :::||||||      |||:::
1242 laThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThr 1258

3827 TCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGC 3876
     :::||||||||||:::   :::|||:::::::::::::|||         ||
1259 GluAlaGlySerSerIleThrSerThrLysGlyGlnValAspLeuLeuAl 1275

3877 ANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTA 3926
     |      |||::::::|||   |||   |||      |||:::   :::|
1275 a......GlnAsnGlySerIleAlaGlySerIleAsnAlaAlaAsnValT 1290

3927 CGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAAC 3976
     ||               |||      :::   |||   |||      :::
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAspIleLys 1306

3977 CGTACATCGGCGCAACGCGCTATTTCGTCCAAA 4009
        ||||||:::      :::|||::::::|||
1307 AlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732

seq_documentation_block:

ID AAR41732 standard; Protein; 1529 AA.
XX
AC AAR41732;
XX
DT 26-APR-1994 (first entry)
XX
DE High molecular weight protein 4 (HMW4).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus influenzae.
XX
PN WO9319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Barenkamp SJ;
XX
DR WPI; 1993-320683/40.
DR N-PSDB; AAQ49511.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX
PS Claim 6; Figure 10; 100pp; English.

XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
SQ Sequence 1529 AA;

alignment_scores:
Quality: 258.00 Length: 1121
Ratio: 0.484 Gaps: 51
Percent Similarity: 47.547 Percent Identity: 20.428

alignment_block:
US-09-303-518D-651 x AAR41732 ..

Align seq 1/1 to: AAR41732 from: 1 to: 1529

```
 99 GTTCGGCATTCTTCCCCAAGCTTGGGCGGGACACACTTATTTCGGCATCA 148
616 LeuArgSerSerArgArgSerPheAlaGlyValHis.PheAsnGlyIleG 632

149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
632 ly.........................GlyLysThrAsnPheAsnIle 639

199 GGGGCGAAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
640 GlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThrAs 656

249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACG 298
656 pProLysLysGluLeuProIle.....ThrPheAsnAlaAsnI 669

299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
669 leThrAlaThrGlyAsnSerAspSerSerValMetPheAspIleHis... 684

349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCA 398
685 .........AlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSe 698

399 GCACCGTTTTTCT.....TACCAAATTGTGAAAAGAAATAATT 436
698 rIleAsnIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArg. 714

437 ATAAGCCTGACAATTCACACCCTTACAACGGCGATTANCATATGCCGCGT 486
715 ...........AsnSerAsnAlaPhe.....Glu 720

487 TTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACAT 536
721 IleLysLys.....AspLeuThrIleAsnAl 729

537 GAGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCA 586
729 aThrGlySerAsnPheSerLeuLysGlnThr..... 739

587 TCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAACACGGCGAT 636
740 .....LysAspSerPheTyrAsnGluTyr......SerLysHisAlaIle 752

637 TTATCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATATGCAGGG 686
753 AsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThrLeu..... 767

687 TTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCA 736
768 .GlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleThrA 784

737 ACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTTCG 786

784 snLysAlaAsn.....ValThrLeuGlnAlaAspThrSerAsnSer 797

787 CCAATGTTTATTTATGACAAAACA.....AACAATAAATGGCTGCT 827
798 AsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValGl 814

828 CAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTT 877
814 uGlyAsnLeuSerLeuThrGly.....AlaAsnAlaA 825

878 TCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGAT 927
825 snIleValGlyAsnLeuSerIleAlaGluAspSerThrPheLysGlyGlu 841

928 ACACATACCGTCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTAC 977
842 Ala.....Se 843

978 ATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNT 1027
843 rAspAsnLeuAsnIleThrGlyThrPhe.....ThrAsnAsnGlyThrA 858

1028 CCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTG 1077
858 laAsnIleAsnIleLysGlnGlyValValLysLeuGlnGlyAspIleIle 874

1078 AATGAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTA 1127
875 AsnLys.....GlyGly..... 878

1128 CCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCA 1177
879 .....LeuAsnIleThrThrAsnAlaSer.....GlyT 888

1178 ACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTG 1227
888 hrGlnLysThrIleIleAsnGlyAsnIleThrAsnGluLysGlyAspLeu 904

1228 TATTTT..... 1233
905 AsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyGlyAsnIl 921

1234 .....GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGT 1273
921 eSerGlnLysGluGlyAsnLeuThrIleSerSerAspLys..... 934

1274 GGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAA 1323
935 .....ValAsnIleThrAsnGlnIleThrIleLysAlaGly 946

1324 GTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCT 1373
947 ValGluGlyGlyArgSerAspSer...SerGluAlaGluAsnAlaAsnLe 962

1374 GCACGTTCAAGCCAAA.....G 1390
962 uThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerG 979

1391 GGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 1440
979 lyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeu..... 993

1441 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTT 1490
994 .....ThrIleGlyAs 997

1491 GNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT...AATCAGTTCA 1537
997 nAlaSerGlyGly.....AsnAlaAspAlaLysLysValT 1009

1538 ACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAAC 1587
```

```
1009 hrPheAspLysVal.....LysAspSerLysIleSerThrAsp 1021

1588 GGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGAT 1637
     ||||||:::::::::::::::          :::|||:::::::|||::: 
1022 GlyHisAsnValThrLeuAsnSerGluValLysThrSerAsnGlySerSe 1038

1638 GATTGNCNATCATAATGCCACAACAACATCCACCGTTACC.....A 1678
        :::   ::::::   :::|||      |||::::::        : 
1038 rAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAlaLysAspV 1055

1679 TTACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATCAAT... 1725
     ::|||   |||:::::::::|||      |||   |||:::|||||| 
1055 alThrValAsnAsnAsnValThr......SerHisLysThrIleAsnIle 1069

1726 ..............AGACTTAATTACAGCAAAGAAATTGCCTACAACGG 1760
                    :::::::::   ::::::      :::   |||:: 
1070 SerAlaAlaAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAl 1086

1761 TTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTG 1810
     :      |||:::   ::::::|||      ||||||
1086 aThrThrGlySerValGluValThrAlaGlnAsnGly............. 1098

1811 TTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACA 1860

1098 ..................................................... 1098

1861 AATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGG 1910
     :::::::::||||||||||||::   |||
1099 ThrIleLysGlyAsnIleThrSerGlnAsn................... 1108

1911 CAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAA 1960
           |||      |||   :::::|||
1109 ....ValThrValThrAlaThrGluAsnLeu.................. 1117

1961 TGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNAC 2010
                          ::::::   :::|||      |||
1118 ......................ValThrThrGluAsnAlaValIleAsn 1126

2011 CGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGAT 2060
        |||      :::   |||:::      :::   |||:::      ||
1127 AlaThrSerGlyThrValAsnIleSerThrLysThrGlyAsp......Il 1141

2061 TTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACG 2110
     |:::   ::::::   :::::::::|||:::   :::::::::      :
1141 eLysGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerG 1158

2111 CCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACA 2160
     ::::::::::      |||:::         ::::::   :::   :::
1158 lyAsnThrLeu...LysValSerAsnIleThrGlyGlnAspValThrVal 1173

2161 CGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGA 2210
        :::|||      :::|||||||:::   :::   ::::::|||:::
1174 ThrAlaAspAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAl 1190

2211 CGATAAAGTGATTGCTTCATTG...ACTAAGACN...GACNTNAGCGGCA 2254
                  |||::::::   |||||||||   |||   :::|||
1190 aThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyL 1207

2255 NTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCA 2304
        |||      ::::::   :::   :::   ::::::      |||:::
1207 ysValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThr 1223

2305 CTNAAN...GGCAATCTTAGTGCAAATGGCGATACA.............. 2337
     |||:::   ||||||:::|||:::|||      |||
1224 LeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSe 1240

2338 ....CGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCC 2383
         :::   |||   :::   :::::::|||   ||||||:::   |||:
1240 rGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSerV 1257

2384 TCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2433
     :::::   ::::::|||:::          ::::::|||:::::::||||||
1257 alThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsn 1273

2434 NCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACA 2483
     :::                                    |||   :::|||::
1274 Thr.............................ValAlaValThrAlaSe 1280

2484 AAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATT 2533
     ::::|||:::||||||:::::::::::::||||||:::::::::::::::
1280 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnG 1297

2534 CCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTT 2583
     ::|||                                 |||::::::
1297 lyAla.............................AlaThrLeuThrAla 1303

2584 GAAAACAGCCGCTTTACCGGACAA.............CTCAGCGGCAG 2618
     |||:::::::::::::|||   |||                  |||:::::
1304 GluSerGlyLysLeuThrThrGlnThrGlySerSerIleThrSerSerAs 1320

2619 CAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAG 2668
     :      |||:::|||      |||||||||
1320 nGlyGlnThrThrLeuThrAlaLysAspSer.................. 1330

2669 GCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAAT 2718
       :::      ||||||:::|||        |||::::::|||||||||
1331 ..SerIleAlaGlyAsnIleAsnAla......AlaAsnValThrLeuAsn 1344

2719 TCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGA 2768
     :::                                  |||||||:::
1345 Thr.............................ThrGlyThr........ 1348

2769 CACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAA 2818
                                        |||:::::::|||      :
1349 .............................LeuThrThrThrGlyAspS 1355

2819 CTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAAC 2868
     :::::::::::::      :::||||||||||:::|||:::|||
1355 erLysIleAsnAlaThrSerGlyThrLeuThrIleAsnAlaLys...... 1369

2869 NGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGA 2918
                                                  :::
1370 .............................................AspAl 1371

2919 CAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCA 2968
       ||||||      |||   |||:::::::   |||   :::   |||:::|
1371 aLysLeuAspGlyAlaAlaSerGlyAspArgThrValValAsnAlaThrA 1388

2969 ACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAA 3018
     ||   :::|||:::      |||:::            :::   |||:::
1388 snAlaSerGlySerGlyAsnValThrAlaLysThrSerSerValAsn 1404

3019 GGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAA 3068
            :::          ::::::||||||:::      :::||
1405 IleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluAs 1421

3069 CGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACG 3118
     |               |||           ::::|||      |||:::
1421 n..............GlyArgAsnThrValArgLeuArgGlyLysGluI 1433

3119 GCGAGTTCCGCCTGCATAATCCG..................... 3141
       :::   :::      :::|||
1433 leAspValLysTyrIleGlnProGlyValAlaSerValGluGluValIle 1449

3142 .................GTCAAAGAACAAGAGCTTTCCGAC........ 3165
                        |||   ::::::|||||||||
1450 GluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGluAr 1466

3166 ............AAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAA 3202
                 |||||||||   :::::::|||   :::   :::|||
1466 gGluThrLeuAlaLysLeuGly...ValSerAlaValArgPheValGluP 1482

3203 AAGACAACGCG 3213
       :::||||||
1482 roAsnAsnAla 1485
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609

seq_documentation_block:
ID   AAY81609 standard; Protein; 1237 AA.
XX
AC   AAY81609;
XX
DT   24-MAY-2000  (first entry)
XX
DE   Streptococcus pneumoniae type 4 protein sequence #109.
XX
KW   Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW   antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW   pneumococcal disease.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006737-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02451.
XX
PR   27-JUL-1998;   98GB-0016337.
PR   19-MAR-1999;   99US-0125164.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Gilbert CFG,  Hansbro PM;
XX
DR   WPI; 2000-195300/17.
XX
PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
PT   pneumococcal diseases and for screening agents capable of antagonizing
PT   or inhibiting expression of the protein  -
XX
PS   Claim 1; Page 83; 108pp; English.
XX
CC   AAY81501 to AAY81679 represent specifically claimed protein sequences
CC   isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC   specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC   The sequences have antibacterial and antiinflammatory properties.
CC   The protein sequences, and fragments of them, are useful as immunogens
CC   and/or antigens. The nucleotide sequences can be used in vaccines and in
CC   diagnostic assays. The proteins and nucleotides can be useful for the
CC   detection and diagnosis of S. pneumoniae. The protein sequences are also
CC   useful for screening an agent capable of antagonising, inhibiting or
CC   interfering with the function or expression of the proteins in which the
CC   agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC   and meningitis. AAA05591 to AAA05614 represent primers used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   1237 AA;


alignment_scores:
            Quality:  257.50            Length:   1198
              Ratio:  0.441               Gaps:     43
 Percent Similarity:  48.748   Percent Identity: 19.699

alignment_block:
 US-09-303-518D-651 x AAY81609   ..

 Align seg 1/1  to: AAY81609  from: 1  to: 1237

    1100 TTTACGCGGCAG...GGGGTGTTAATCAGTACCGTCCAA.........GG 1137
         ||||||:::::   :::||||||   :::|||||||:::         ::
      40 PheThrAsnGluSerAlaValLeuGluLysThrValGluLysThrAspAl 56

    1138 TTAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACT 1187
         :   :::|||      |||   :::|||:::      :::|||:::||||
      56 aLeuAlaThrAsnAspThrValValLeuGlyThrIleSerThrSerAsnS 73

    1188 CATCTTATCAAACAACA.....................TCAACCAAGGCG 1216
         ||:::   :::|||:::                     ||||||:::|||
      73 erAlaSerSerThrSerLeuSerAlaSerGluSerAlaSerThrSerAla 89

    1217 CGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAAAACAAC 1266
               :::      ::::::         :::      ::::::::
      90 SerGluSerAlaSerThrSerAlaSerThrSerAlaSerThrSerAlaSe 106

    1267 GAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAG...ACAGTACCGT 1313
         ::::   :::::::::   |||      |||::::::   :::      :
     106 rGluSerAlaSerThrSerAlaSerThrSerIleSerAlaSerSerThrV 123

    1314 TACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCA 1363
         :::::|||:::   ||||||      |||   |||      |||::::::
     123 alValGlySerGlnThrAlaAlaAlaThrGluAlaThrAlaLysLysVal 139

    1364 AAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1413
         :::   :::         ||||||   :::         |||      ::
     140 Glu...GluAspArgLysLysProAlaSerAspTyrValAlaSerValTh 155

    1414 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAA 1463
         :   :::::::::|||||||:::         |||:::            :
     155 rAsnValAsnLeuGlnSerTyrAlaLysArgArgLysArgSerValAspS 172

    1464 AAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGC 1513
         ::   :::   |||:::   |||      :::||||||
     172 erIleGluGlnLeuLeuAlaSerIleLysAsnAlaAlaValPheSerGly 188

    1514 AACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC... 1560
         |||         |||:::::::::::::|||:::   :::   |||
     189 Asn......ThrIleValAsnGlyAlaProAlaIleAsnAlaSerLeuAs 203

    1561 .........................TTTCGCGGCGGACGTTTGGATTTAA 1585
                                   :::   |||      :::|||
     203 nIleAlaLysSerGluThrLysValTyrThrGlyGluGlyValAspSerV 220

    1586 ACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCG 1635
               :::          :::::::::::::::   |||:::::::|||:::
     220 alTyrArgValProIleTyrTyrLysLeuLysValThrAsnAspGlySer 236

    1636 ATGATT.......................................... 1641
            :::
     237 LysLeuThrPheThrTyrThrValThrTyrValAsnProLysThrAsnAs 253

    1642 ................................GNCNATCATAATGCCA 1657
                                        :::   :::|||:::
     253 pLeuGlyAsnIleSerSerMetArgProGlyTyrSerIleTyrAsnSerG 270

    1658 CAACAACATCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCG 1707
           |||:::::::   :::   :::|||   ::::::::::::   :::|||
     270 lyThrSerThrGlnThrMetLeuThrLeuGlySerAspLeuGlyLysPro 286

    1708 AGTGGTAAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAA 1757
         ||||||   :::            ||||||
     287 SerGlyValLys............AsnTyr..................... 292

    1758 CGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACC 1807
                                     |||      |||||||||      |
     293 .....................IleThrAspLysAsnGlyArgGlnValL 302

    1808 TTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGA 1857
         ||   |||:::   :::            |||         :::|||:::
     302 euSerTyrAsnThrSer............ThrMetThrThrGlnGlySer 314

    1858 ACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAG 1907
                  ||||||   :::|||   ||||||      ||||||::
     315 GlyTyrThrTrpGlyAsnGlyAlaGlnMetAsnGly......PhePheAl 329

    1908 CGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAA 1957
         :   :::              :::|||:::   |||   |||:::|||:::
     329 aLysLys............GlyTyrGly...LeuThrSerSerTrpThr. 340

    1958 AAATGGAAGGTATCCCA...CAAGGAGAAATCGTGTGGGACAACGACTGG 2004
```



```
3581 TGTTTGCCGAAGACCG.....................CCGCAACGCNGTT 3609
          |||:::|||                       ||||||:::   |
 934 lProValArgGlnProGlnArgArgGlnValArgGlnProGlnGlnValL 951

3610 TG..................GACAAGCNGCAT..............CCG 3626
     ||                  ::::::   |||              |||
 951 euValHisGlnLeuGlnHisGlnArgValHisArgLeuArgArgGlnPro 967

3627 GNACACCAAACA.....................CTACCGTTCGCAAGATT 3655
        ||||||:::                    ||||||      :::
 968 ValHisGlnSerGlnGlnValProValArgGlnLeuProHisGlnGlnVa 984

3656 TCCGCGC.................................CTACCGC 3669
     ||||||                                 |||   |
 984 lProArgLeuGlnGlnAlaProValArgArgLeuGlnGlnValLeuAlaP 1001

3670 CAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGG 3719
     ||      :::|||   |||   |||      ::::::      ||||||
1001 roGlnProGlnProGlnProValArgGlnProGlnGlnValSerGlnArg 1017

3720 GCGCGTCGGCATCCT.....................GTTTTCGCACA 3745
        :::||||||                       :::      ||
1018 LeuAsnArgHisGlnArgValArgProLeuGlnGlnValLeuAlaProGl 1034

3746 ACCGGACCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGCTT 3795
     ||||:::|||:::   :::::::|||   :::||||||
1034 nProGlnArgGlnGlnValHisArgLeuGlnArgGln............. 1046

3796 GCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGA..... 3839
               |||   |||      ||||||::::::||||||
1047 ...........ArgValArgLeuAsnArgHisGlnArgValArgProLeu 1059

3840 .....................................CATCGGCATCA 3850
                                         ||||||   ||
1060 GlnGlnValLeuAlaProGlnProGlnArgGlnGlnValHisArgLeuGl 1076

3851 GCACGGGCGCGGG..................................... 3863
     ||||   |||
1076 nHisGlnArgValArgProLeuGlnGlnValLeuAlaProGlnProGlnA 1093

3864 ...................TTTTAGCAGCGGCANTCTNTCAGACGGCAT 3893
                         ||||||             :::|||
1093 rgGlnGlnValHisArgLeuGlnArgGlnArgValArgLeuSerGlnHis 1109

3894 CGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGAT 3943
     :::|||   :::   |||   :::|||
1110 GlnArgValArgGlnProGlnGlnAla..................... 1118

3944 ACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACG 3993
                        |||:::   ::::::   ::::::
1119 .......................HisGlnLeuLeuAsnLeuHisGlnP 1127

3994 CGCTATTTCGTCCAAAA.....................AGCGGATTACCG 4022
        :::   :::|||:::                   :::   :::|||
1127 roValArgGlnProGlnHisArgGlnAlaProGlnLeuGlnGlnValPro 1143

4023 CTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAA 4064
     :::|||:::   |||   |||:::   |||         |||
1144 ValArgGlnProGlnArgArgGlnValArgArgLeuGlnGln 1157
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292

seq_documentation_block:
ID   AAW30292 standard; Protein; 1601 AA.
XX
AC   AAW30292;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW4; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 372
FT                   /note= "encoded by TCT"
FT   Misc-difference 400
FT                   /note= "encoded by AAT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90993.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 97-102; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW4 amino acid sequence was deduced from an
CC   isolated hmw4 gene (see AAT90993).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW3 (see AAW30291) have also been identified.  A
CC   conjugate comprising HMW4 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW4 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1601 AA;


alignment_scores:
           Quality:  256.50              Length:   1481
             Ratio:  0.374                 Gaps:     73
 Percent Similarity:  46.253   Percent Identity: 20.257

alignment_block:
 US-09-303-518D-651 x AAW30292   ..

 Align seg 1/1  to: AAW30292  from: 1  to: 1601

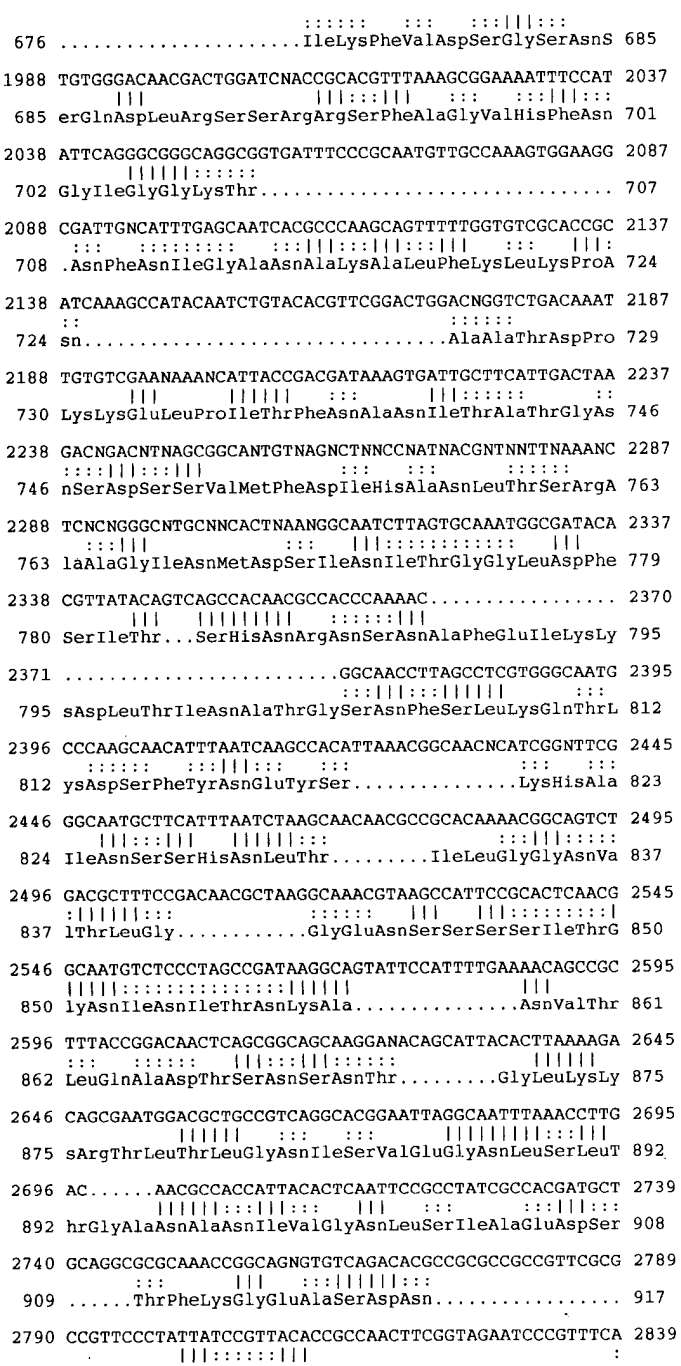
     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
         :::|||         |||   :::   |||:::::::::::   |||||
     158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

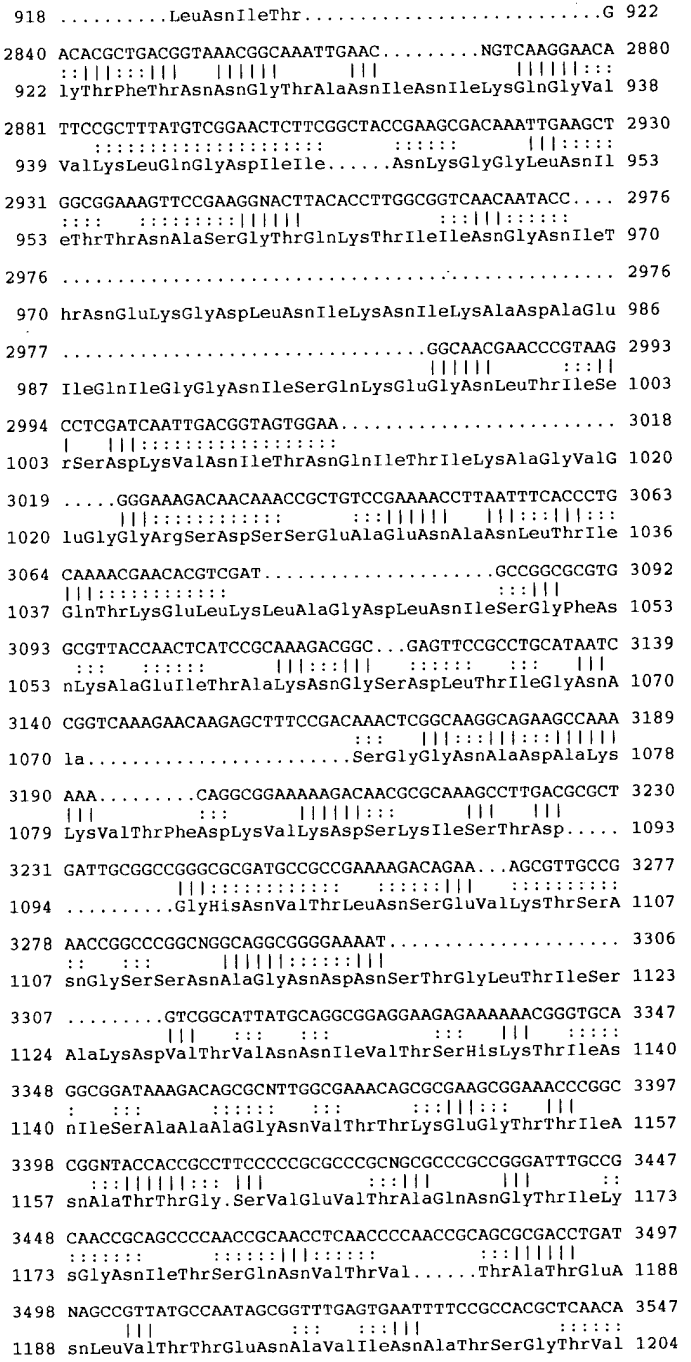
     213 TGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
         |:::      |||   |||::::::          ::::::||||||
     174 eSerAsnGluAsnIleLysAlaArgAsnPheThrLeuGluGlnThrLysA 191

     262 ...........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                 |||   :::::::::   ::::::      |||::::::
     191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

     295 AACGGCGTGGCGGCATTGGTGGGC............GATCAATATATTGT 332
         :::|||   :::   |||:::|||           ::::::   :::::
     208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224
```




676 IleLysPheValAspSerGlySerAsnS 685
1988 TGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTCCAT 2037
685 erGlnAspLeuArgSerSerArgArgSerPheAlaGlyValHisPheAsn 701
2038 ATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGG 2087
702 GlyIleGlyGlyLysThr 707
2088 CGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGC 2137
708 .AsnPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProA 724
2138 ATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAAT 2187
724 sn.AlaAlaThrAspPro 729
2188 TGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGACTAA 2237
730 LysLysGluLeuProIleThrPheAsnAlaAsnIleThrAlaThrGlyAs 746
2238 GACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNAAANC 2287
746 nSerAspSerSerValMetPheAspIleHisAlaAsnLeuThrSerArgA 763
2288 TCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCAAATGGCGATACA 2337
763 laAlaGlyIleAsnMetAspSerIleAsnIleThrGlyGlyLeuAspPhe 779
2338 CGTTATACAGTCAGCCACAACGCCACCCAAAAC 2370
780 SerIleThr...SerHisAsnArgAsnSerAsnAlaPheGluIleLysLy 795
2371 .GGCAACCTTAGCCTCGTGGGCAATG 2395
795 sAspLeuThrIleAsnAlaThrGlySerAsnPheSerLeuLysGlnThrL 812
2396 CCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACNCATCGGNTTCG 2445
812 ysAspSerPheTyrAsnGluTyrSer.LysHisAla 823
2446 GGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGTCT 2495
824 IleAsnSerSerHisAsnLeuThr.IleLeuGlyGlyAsnVa 837
2496 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2545
837 lThrLeuGly.GlyGluAsnSerSerSerSerIleThrG 850
2546 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGC 2595
850 lyAsnIleAsnIleThrAsnLysAla.AsnValThr 861
2596 TTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGA 2645
862 LeuGlnAlaAspThrSerAsnSerAsnThr.GlyLeuLysLy 875
2646 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2695
875 sArgThrLeuThrLeuGlyAsnIleSerValGluGlyAsnLeuSerLeuT 892
2696 AC.AACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCT 2739
892 hrGlyAlaAsnAlaAsnIleValGlyAsnLeuSerIleAlaGluAspSer 908
2740 GCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCG 2789
909 ThrPheLysGlyGluAlaSerAspAsn 917
2790 CCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCA 2839


918 .LeuAsnIleThr.G 922
2840 ACACGCTGACGGTAAACGGCAAATTGAAC.NGTCAAGGAACA 2880
922 lyThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlnGlyVal 938
2881 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCT 2930
939 ValLysLeuGlnGlyAspIleIle.AsnLysGlyGlyLeuAsnIl 953
2931 GGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCAACAATACC.... 2976
953 eThrThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsnIleT 970
2976 2976
970 hrAsnGluLysGlyAspLeuAsnIleLysAsnIleLysAlaAspAlaGlu 986
2977 .GGCAACGAACCCGTAAG 2993
987 IleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSe 1003
2994 CCTCGATCAATTGACGGTAGTGGAA. 3018
1003 rSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGlyValG 1020
3019 GGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTG 3063
1020 luGlyGlyArgSerAspSerSerGluAlaGluAsnAlaAsnLeuThrIle 1036
3064 CAAAACGAACACGTCGAT.GCCGGCGCGTG 3092
1037 GlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerGlyPheAs 1053
3093 GCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAATC 3139
1053 nLysAlaGluIleThrAlaLysAsnGlySerAspLeuThrIleGlyAsnA 1070
3140 CGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAA 3189
1070 la.SerGlyGlyAsnAlaAspAlaLys 1078
3190 AAA.CAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCT 3230
1079 LysValThrPheAspLysValLysAspSerLysIleSerThrAsp..... 1093
3231 GATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAA...AGCGTTGCCG 3277
1094 .GlyHisAsnValThrLeuAsnSerGluValLysThrSerA 1107
3278 AACCGGCCCGGCNGGCAGGCGGGGAAAAT. 3306
1107 snGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSer 1123
3307 .GTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCA 3347
1124 AlaLysAspValThrValAsnAsnIleValThrSerHisLysThrIleAs 1140
3348 GGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGC 3397
1140 nIleSerAlaAlaAlaGlyAsnValThrThrLysGluGlyThrThrIleA 1157
3398 CGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCG 3447
1157 snAlaThrThrGly.SerValGluValThrAlaGlnAsnGlyThrIleLy 1173
3448 CAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGAT 3497
1173 sGlyAsnIleThrSerGlnAsnValThrVal.....ThrAlaThrGluA 1188
3498 NAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACA 3547
1188 snLeuValThrThrGluAsnAlaValIleAsnAlaThrSerGlyThrVal 1204

```
     3548 GCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGC 3597
          :::|||                                    ||||||::
     1205 AsnIleSer.............................ThrLysThrGl 1211

     3598 CGCAACGCNGTTTGGA...CAAGCNGCATCCGGNACACCAAACACTACCG 3644
          :          |||   ::::::::::||||||:::   |||   |||:
     1211 yAspIleLysGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrA 1228

     3645 TTCGC............AAGATTTCCGCGCCTACCGCCAACAAACCGACC 3682
          ::              |||:::|||      |||::::::       |||
     1228 laSerGlyAsnThrLeuLysValSerAsnIleThrGlyGlnAspValThr 1244

     3683 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATC 3732
             :::   ::::::          |||:::::::||||||::::::   ||
     1245 ValThrAlaAspAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSe 1261

     3733 CTGTTTTCGCACAACCGGA.............................. 3751
          |:::      |||||||||
     1261 rAla......ThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspI 1276

     3752 .................CCGAAAACANCTTCGACGACGGCATCGGCAACT 3784
                           :::::::   |||:::|||      ||||||
     1276 leAsnGlyLysValGluSerSerSerGlySerValThrLeuValAlaThr 1292

     3785 CGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT............AC 3822
                |||   ::::::   :::|||||||||              .||
     1293 GlyAlaThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleTh 1309

     3823 GGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAG 3872
          ||||:::::::|||:::   ::::::::::::   :::   :::   ::::
     1309 rAlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyT 1326

     3873 CGGCANTCTNTCAGACGGCATCGGAGGCAAAA................... 3904
          ::   :::      |||:::|||::::::
     1326 hrAsnSerValThrThrSerSerGlnSerGlyAspIleGluGlyThrIle 1342

     3905 TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGAT...ACCGCGCC 3951
          |||:::   :::      :::||||||          |||   |||
     1343 SerGlyAsnThrValAsnValThrAlaSerThrGlyAspLeuThrIleGl 1359

     3952 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTT 4001
             ||||||          :::::::::   :::|||      :::  :
     1359 yAsnSerAlaLysValGluAlaLysAsnGlyAlaAlaThrLeuThrAlaG 1376

     4002 CGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG 4042
          ::|||   |||      |||   |||   :::|||||||:::
     1376 luSerGlyLysLeuThrThrGlnThrGlySerSerIleThr 1389
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01847

seq_documentation_block:
ID   AAB01847 standard; Protein; 1095 AA.
XX
AC   AAB01847;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 mature HMW1A protein, SEQ ID NO:69.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52196.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 28F-Q; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1095 AA;


alignment_scores:
           Quality:  254.50              Length:   1062
             Ratio:  0.474                 Gaps:     52
 Percent Similarity:  50.565   Percent Identity: 20.433

alignment_block:
 US-09-303-518D-651 x AAB01847   ..

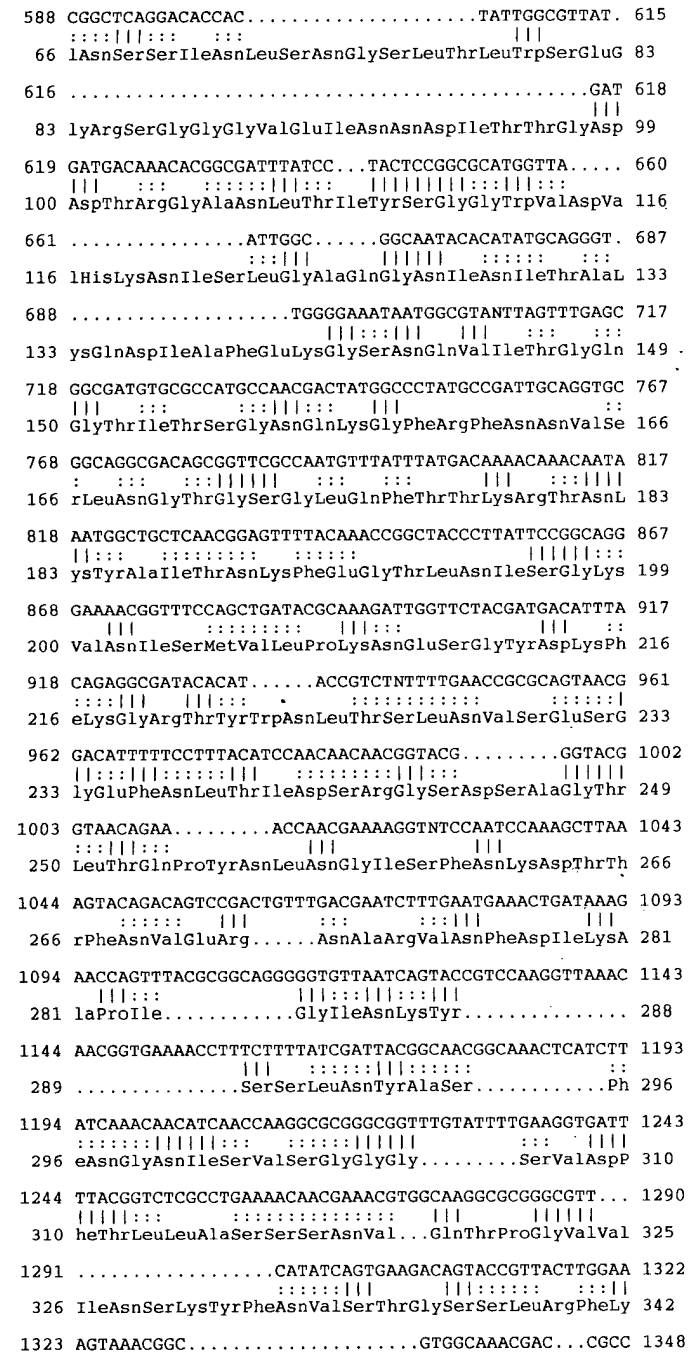
 Align seg 1/1  to: AAB01847  from: 1  to: 1095

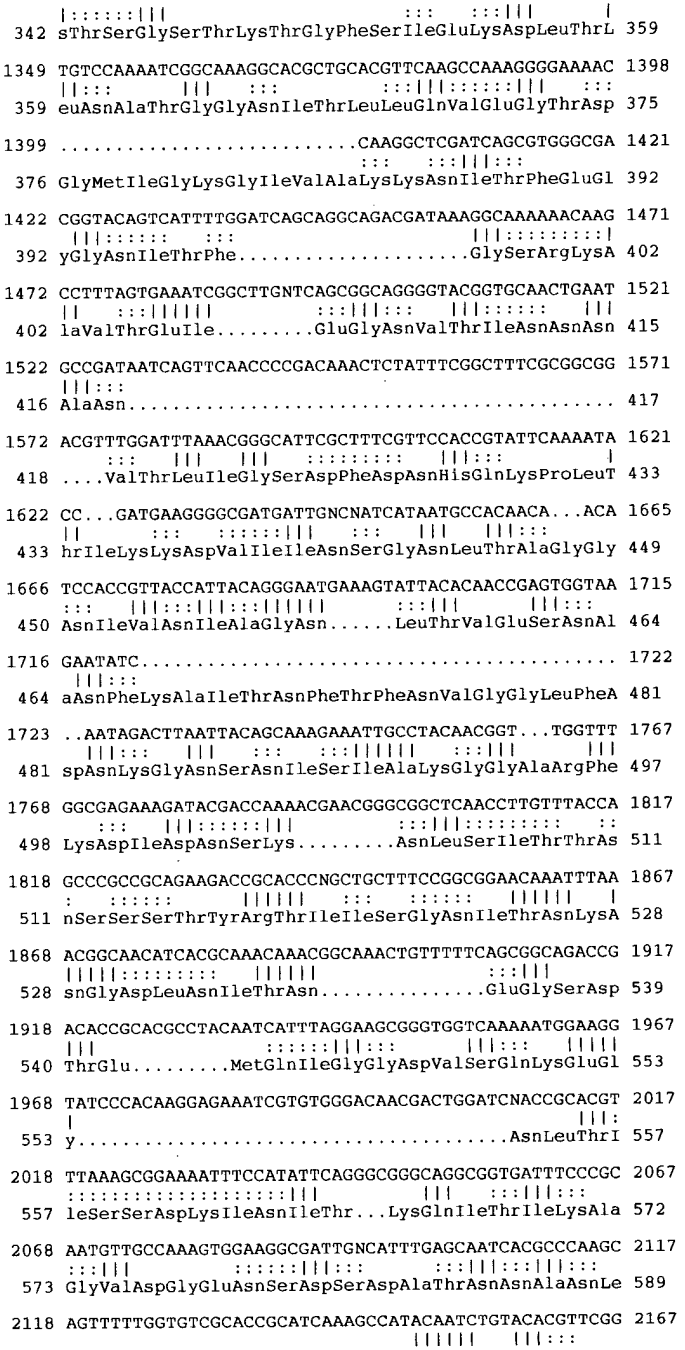
     391 CCCGATCAGCACCGTTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
         ||||||:::      ::::::   :::   :::   |||:::|||   ::
       1 ProAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSe 17

     441 GCCTGACAATTCACACCCTTACAACGGCGAT...TANCATATGCCGCGTT 487
         :   |||::::::::      :::|||:::   :::      |||:::
      17 rGluAspAspGluTyrThrGlySerGlyAsnSerAlaSerThrProLysA 34

     488 TGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 537
           :::|||      |||           :::   :::   |||   :::
      34 rgAsnLysGluLysThrThrLeuThrAsnThrThrLeuGluSerIleLeu 50

     538 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCAT 587
         :::   :::|||:::   :::               :::|||:::   ::
      51 LysLysGlyThrPheValAsnIleThrAla...AsnGlnArgIleTyrVa 66
```





```
 589 u..........................ThrIleLysThrLys.... 594

2168 ACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAA 2217
                              |||      :::|||:::|||
 595 ........................GluLeuLysLeuThrGlnAspLeu 602

2218 GTGATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNC 2267
     |||:::::::::::::|||::::::   ::::::
 603 AsnIleSerGlyPheAsnLysAlaGluIleThrAlaLys..... 615

2268 CNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATC 2317
                             |||:::              ||||||
 616 .....................AspGlySerAspLeuThrIleGlyAsnT 625

2318 TT...AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC 2364
       ||||||:::|||   :::          |||:::   |||   :::
 625 hrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVal 641

2365 CAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCA 2414
     ::::::::::::::|||      |||                  :::::
 642 LysAspSerLysIleSerAlaAspGly.................HisLy 652

2415 AGCCACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTAATC 2464
     ::::|||||||:::::::::   ::::::||||||:::   :::   |||
 652 sValThrLeuHisSerLysValGluThrSerGlySerAsnAsn...AsnT 668

2465 TAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCT 2514
       ::::::::::::   :::|||::::::||||||:::       :::|||
 668 hrGluAspSerSerAspAsnAsnAlaGlyLeuThrIle......AspAla 682

2515 AAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGA 2564
     |||   ||||||:::          :::|||:::|||::::::   :::
 683 Lys...AsnValThr.........ValAsnAsnAsnIleThr...SerHi 694

2565 TAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGCG 2614
      |||||||||          :::   |||      :::|||:::::::::
 694 sLysAlaVal.........SerIleSerAlaThrSerGlyGluIleThrT 708

2615 GCAGCAAGGANACAGCATTACACTTAAAA......GACAGCGAATGGACG 2658
       :::      |||:::::::::             :::   ||| .  |||
 708 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 724

2659 CTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCAT 2708
           :::|||:::   ||||||::::::::   ::::::::::::::
 725 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 741

2709 TACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCA 2758
     :||||||::::::::::          :::::::|||   |||   ::::::
 741 lThrLeuThrAlaThr.........GluGlyAlaLeuAlaValSerAsnI 755

2759 GNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTT 2808
       :::::::|||             ::::::   :::   |||::::::
 755 leSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuThrThr 771

2809 ACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGG 2858
            :::::::::::::::      ::::::   |||   :::
 772 LeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerSerGl 788

2859 CAAATTGAACNGTCAAGGAACATTC.....................CGCT 2887
      :::   :::     ||||||:::                      :::
 788 nSerGlyAspIleGlyGlyThrIleSerGlyGlyThrValGluValLysA 805

2888 TTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAA 2937
       ::::::|||             ::::::|||:::|||   |||
 805 laThrGluSerLeuThrThrGlnSerAsnSerLysIleLys...AlaThr 820

2938 AGTTCCGAAGGNACTTACACCTTGGCGGTC..................... 2967
     ::::::|||::::::   |||   |||:::
 821 ThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGlyThrIl 837

2968 ....AACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTA. 3012
       :::||||||   |||      ::::::      :::||||||||||
 837 eSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrValG 854

3013 .....................GTGGAAGGGAAAGACAACAAACCGCTG 3039
                                 :::||||||      :::
 854 lyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrThr 870

3040 TCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGC 3089
     |||::::::         |||      :::::::|||:::   ::: |||:
 871 SerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrSer.AlaL 887

3090 GTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATC 3139
     ::|||   :::|||   |||||||:::   :::|||   |||      |||
 887 ysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGlySerIle 903

3140 CGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAA 3189
     ::::::   |||   ::::::   |||:::::::::            ||
 904 AsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValLy 920

3190 AAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3223
     |:::   :::   ::::::|||      :::|||
 920 sGlySerAsnIleAsnAlaThrSerGlyThrLeu 931
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01833

seq_documentation_block:

```
ID   AAB01833 standard; Protein; 1005 AA.
XX
AC   AAB01833;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K21.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA000938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52182.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 21A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
```

CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1005 AA;

alignment_scores:
Quality: 252.00 Length: 1177
Ratio: 0.451 Gaps: 53
Percent Similarity: 47.494 Percent Identity: 19.031

alignment_block:
US-09-303-518D-651 x AAB01833 ..

Align seq 1/1 to: AAB01833 from: 1 to: 1005

```
226 AAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGA 275
    :::|||||||      |||:::   |||   ||||||        :::::
 45 GlnLysGlyIleGluValAsnIleSerAlaThrLysAsnValThrValAs 61

276 TTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAAT 325
    :   :::|||      :::|||||:::   :::|||   :::::::::
 61 nAlaAspValAspValLysAsnGlyThrLeuValLeuHisSerGlnArgA 78

326 ATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT 375
        |||::::::      ||||||:::   ::::::::::::   |||
 78 snGlyValLysIle......AsnGlyAsnIleThrSerThrGlnAsnGly 92

376 ............GCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTA 413
                   |||   :::   |||   |||:::
 93 AsnLeuThrIleLysThrGlyGlyLysValAspValHisLysAsnIleTh 109

414 CCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACA 463
        :::         |||         ||||||:::
109 rLeuGlyMetGlyPheLeuAsnIleThrSerAspAsnAsn.......... 122

464 ACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAA 513
                                  :::|||      :::
123 ................................IleThrPheGluLys 127

514 CCTGTCGAAATGACGAGTGACATGAGGGGGAATACCTATTCCGATAAAGA 563
        ::::::|||         :::||||||      |||:::::::||
128 GlyAspAsnLeuThrIleThrAlaGlnGlyAsnIleIleSerAsnGlnGl 144

564 AAAATATCCCGAGCGTGTCCGCATCGGCTCA.................... 594
    |:::      :::::::::|||:::::::::
144 uAsn......LysGlnLeuArgPheSerAsnValSerLeuAsnGlyMetG 159

595 ............................GGACACCACTATTGGCGTTAT 615
                                  |||:::|||      ::::::
159 lyAlaGlyLeuThrPheThrAlaAsnLysGlyAsnHisThrHisLysPhe 175

616 GATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTAATTGG 665
    |||           |||   |||:::   ||||||      :::|||::
176 Asp............GlyThrLeuAsnIleSerGlyLysValValIleAs 188

666 CGGCAATACA......CATATGCAGGGTTGGGGAAATAATGGCGTANTTA 709
    :   :::|||      ::::::      |||:::   :::
188 nGlnThrThrProHisAsnIleAlaProTrpAsnAlaSer.......... 201

710 GTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 759
                              |||::::::|||      :::
202 .....................AlaAspSerTyrTrpAsnValThrThr 210

760 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAAC 809
          ::::::::::::         ::::::   :::      ::
211 LeuThrLeuGlyAsnAsnAlaGlnPheThrPheIleLysPheValAspSe 227

810 AAACAATAAATGGCTGCTCAACGGAGTTTTACAAACC......GGCTACC 853
    :|||::::::      ||||||:::      ::::::      |||
227 rAsnArgSerValAlaLeuAsnSerGlySerArgSerPheAlaGlyValL 244

854 CTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGGTTC 903
      :::   |||::::::|||   ::::::
244 ysPheTyrGlyLysAsnAsnGluMetLys..................... 253

904 TACGATGACATTTACAGAGGCGATACACATACCGTCTNTTTTGAACCGCG 953
          :::      ||||||:::   :::|||   |||:::   ::
254 .........PheAsnIleGlyAspAsnAlaAsnValGluPheLysLeuLy 267

954 CAGTAACGGACATTTTTCCTTTACATCCAACAAC.............. 987
    :||||||   :::         ||||||||||||
267 sSerAsnAspAsn.........ThrSerAsnAsnLysProLeuProIleG 281

988 ......................AACGGTACGGGTACGGTAACA...... 1008
                              :::|||:::||||||||||:::
281 lnPheLeuSerAsnIleSerAlaThrGlyAsnGlyThrValSerPheAsp 297

1009 ...GAAACCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGT 1055
       ::::::|||   ::::::   :::   :::|||:::::::::   ::
298 IleHisAlaAsnLeuSerAlaArgSerThrGluLeuAsnMetSerLeuIl 314

1056 CCGACTG.........TTTGACGAATCTTTGAAT................ 1080
     :::::::            :::   |||:::|||
314 eAsnIleSerAsnGlyValAsnPheSerIleAsnSerHisValArgGlyA 331

1081 ...........GAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTT 1119
                |||       |||:::   :::       |||   |||
331 snAsnAlaPheGluIleLysLysAspLeuIleIleAsnAlaThrGlySer 347

1120 AAT............CAGTACCGTCCAAGGTTAAACAACGGTGAAAACCT 1157
     |||            |||   :::   :::::::::|||
348 AsnPheAsnLeuLysGlnThrLysAspLysPheAspAsn........... 360

1158 TTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACATCA 1207
               :::|||   ::::::   ::::::   :::::::|||:::
361 ..........SerTyrGluLysAsnAlaIlePheSerThrHisAsnLeuT 374

1208 ACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCT 1257
     ::                        :::   |||:::   |||::::::
374 hr..................IleLeuGlyGlyAsnValThrLeuGlyGly 384

1258 GAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAG 1307
     ||||||:::::::::         ::::::::::::|||::::::   ::
385 GluAsnSerSerSerAsnIleLysGlyAsnIleAsnIleAsnSerLysAl 401

1308 TACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAA 1357
     :::::::::   ::::::::::::
401 aAsnValThrLeuGlnAlaHisAla........................ 409

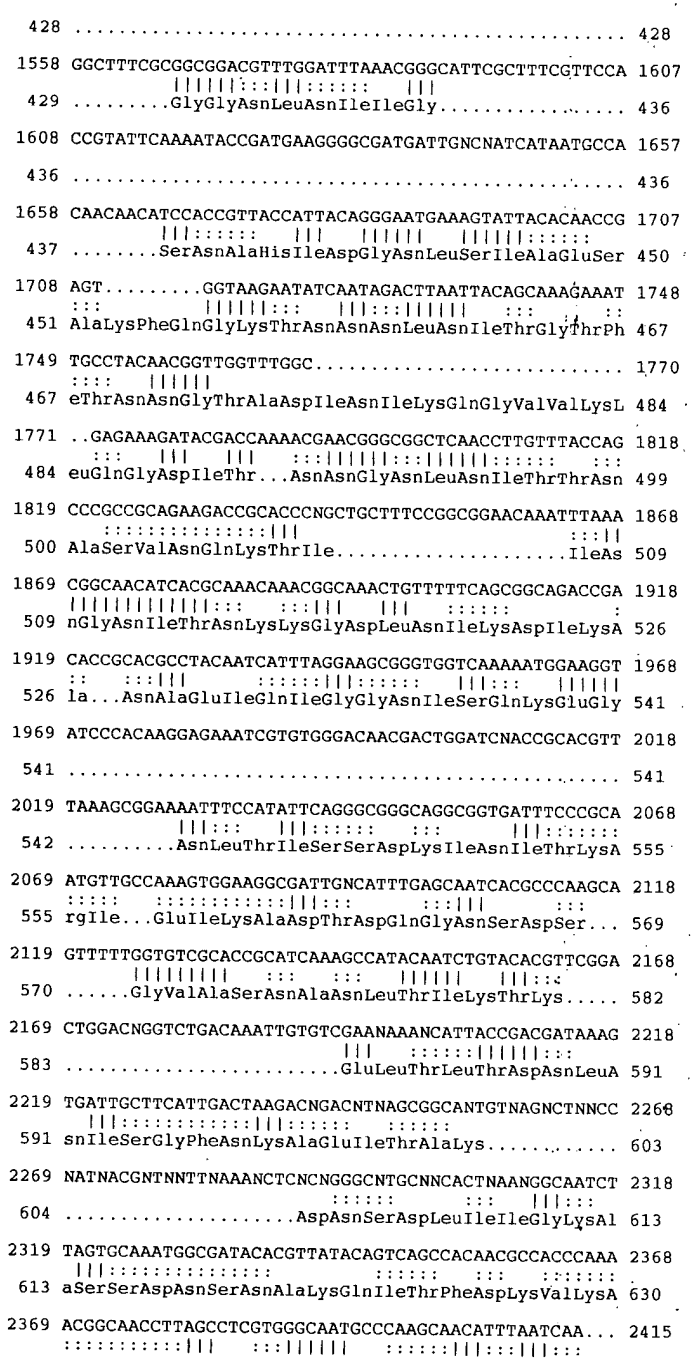
1358 TCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCG 1407
           ||||||   |||::::::   |||          :::   :::
410 ........GlyThrSerHisLeuAspLysLys.........GluArgThr 420

1408 ATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAA 1457
     :::::::::|||:::   :::|||
421 LeuThrLeuGlyAsnValSerVal......................... 428

1458 AGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTA 1507

 428 ............................................... 428

1508 CGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTC 1557
```





```
    3175 AAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3224
           ||||||       :::      :::   ::::::   |||   ::
     906 GlyAlaGluIleAspValLysTyrIleGlnProGlyValAlaSerAlaAs 922

    3225 CGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTG 3274
         :   :::|||   |||   |||   |||   ||||||:::::::::::::
     922 nGluValIleGluAlaLysArg...AlaLeuGluLysValLysAspLeuS 938

    3275 CCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3324
         :::::      |||   :::   ::::::   :::|||:::      :::
     938 erAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal 953

    3325 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAA 3374
                    :::   ::::::   ::::::      ::::::
     954 ............ArgPheIleGluProAsnAsnThrIleThrValAsnTh 966

    3375 ACAGCGCGAAGCGGAAACCCGGCCGGNTACC 3405
          |||:::|||      |||||||||:::::
     966 rGlnAsnGluPheThrThrArgProSerSer 976
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01832

seq_documentation_block:

```
ID   AAB01832 standard; Protein; 1011 AA.
XX
AC   AAB01832;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K21.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52181.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 21A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1011 AA;
```

```
alignment_scores:
           Quality:  252.00              Length:   1177
             Ratio:  0.451                 Gaps:     53
 Percent Similarity:  47.494   Percent Identity: 19.031

alignment_block:
 US-09-303-518D-651 x AAB01832   ..

 Align seg 1/1  to: AAB01832  from: 1  to: 1011

     226 AAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGA 275
         :::||||||      |||:::   |||   ||||||         :::::
      51 GlnLysGlyIleGluValAsnIleSerAlaThrLysAsnValThrValAs 67

     276 TTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAAT 325
         :   :::|||      :::||||||:::   :::|||   :::::::::
      67 nAlaAspValAspValLysAsnGlyThrLeuValLeuHisSerGlnArgA 84

     326 ATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT 375
             |||::::::      ||||||:::   ::::::::::::   |||
      84 snGlyValLysIle......AsnGlyAsnIleThrSerThrGlnAsnGly 98

     376 ............GCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTA 413
                        |||   :::   |||   |||:::
      99 AsnLeuThrIleLysThrGlyGlyLysValAspValHisLysAsnIleTh 115

     414 CCAAATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACA 463
             :::         |||         ||||||:::
     115 rLeuGlyMetGlyPheLeuAsnIleThrSerAspAsnAsn..... 128

     464 ACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAA 513
                                                 :::|||      :::
     129 .................................IleThrPheGluLys 133

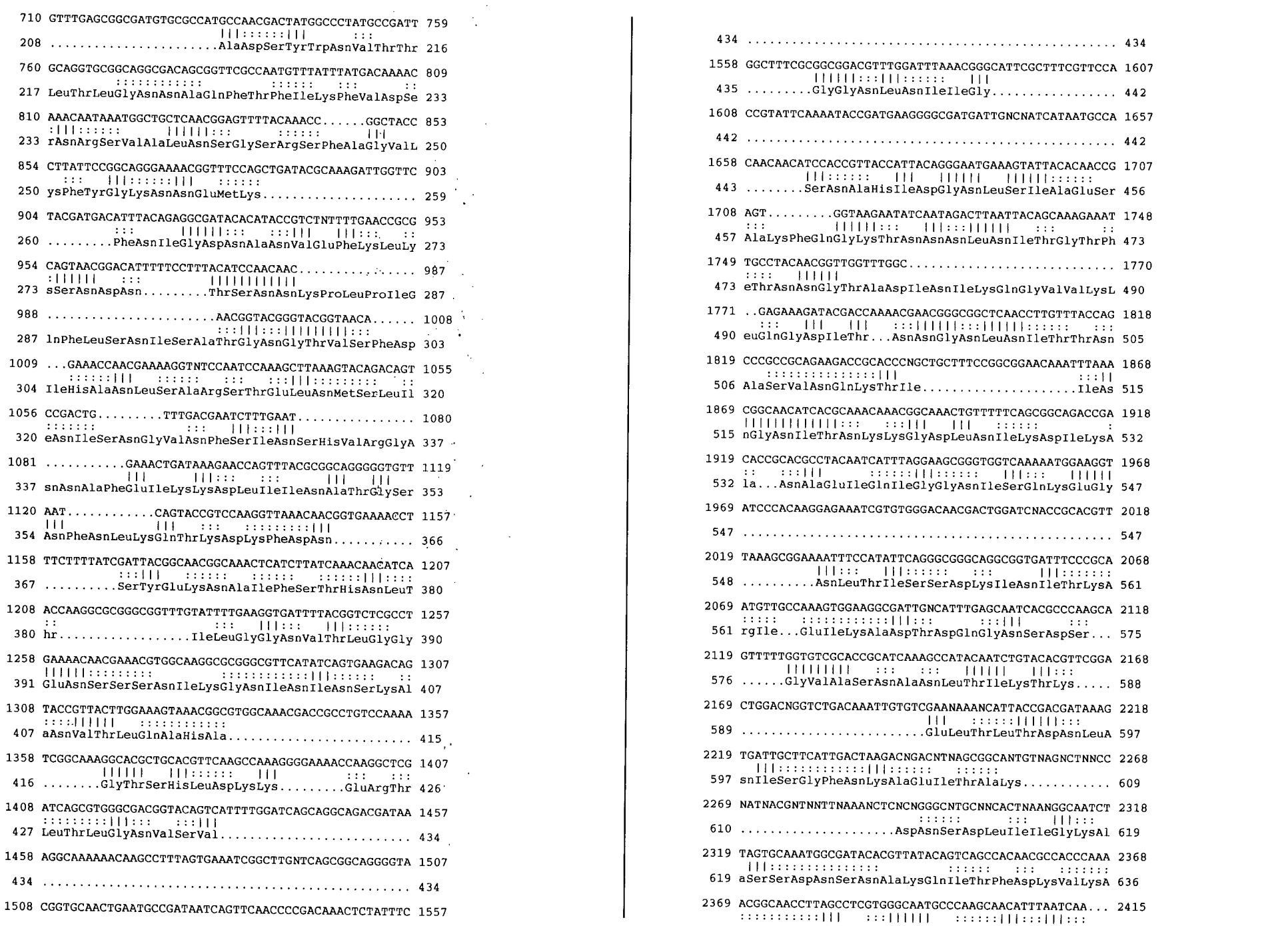
     514 CCTGTCGAAATGACGAGTGACATGAGGGGGAATACCTATTCCGATAAAGA 563
             :::::::|||         :::||||||      |||:::::||
     134 GlyAspAsnLeuThrIleThrAlaGlnGlyAsnIleIleSerAsnGlnGl 150

     564 AAAATATCCCGAGCGTGTCCGCATCGGCTCA.................. 594
         |:::      ::::::::::|||:::::::::
     150 uAsn......LysGlnLeuArgPheSerAsnValSerLeuAsnGlyMetG 165

     595 ............................GGACACCACTATTGGCGTTAT 615
                                       |||:::|||      ::::::
     165 lyAlaGlyLeuThrPheThrAlaAsnLysGlyAsnHisThrHisLysPhe 181

     616 GATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTAATTGG 665
         |||            |||   |||:::   ||||||      :::|||::
     182 Asp............GlyThrLeuAsnIleSerGlyLysValValIleAs 194

     666 CGGCAATACA......CATATGCAGGGTTGGGGAAATAATGGCGTANTTA 709
         :   :::|||      ::::::      |||:::   :::
     194 nGlnThrThrProHisAsnIleAlaProTrpAsnAlaSer..... 207
```



```
 636 spSerLysIleSer...AlaGlyAsnHisAsnValThrLeuAsnSerLys 651

2416 ...GCCACATTAAAC.............GGCAACNCATCGGNTTCGGG 2447,
         |||   |||               ||||||   |||   :::::
 652 ValGluThrSerAsnSerAspGlySerThrGlyAsnGlySerAspAspAs 668

2448 CAATGCTTCATTTAATCTAAGC.......................... 2469
     :|||   ::::::::::::|||
 668 nAsnIleGlyLeuThrIleSerAlaLysAspValThrValAsnSerAsnI 685

2470 ...................AACAACGCCGCACAAAACGGCAGTCTGACG 2499
                         |||   :::|||:::::::|||:::::|||
 685 leThrSerHisLysThrValAsnIleSerAlaSerGluGlyGlyIleThr 701

2500 CTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAA 2549
               :::||||||:::::::::   :::|||   :::|||::
 702 ............ThrLysAlaGlyThrThrIleAsnAlaThrThrGlySe 714

2550 TGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTA 2599
     :|||::::::::::   |||                           |
 714 rValGluValThrAlaLys.............................T 721

2600 CCGGACAACTCAGCGGCAGC.........AAGGANACAGCATTACACTTA 2640
     |||||::::::||||||:::          |||:::::::::::::
 721 hrGlyAspIleSerGlyThrIleSerGlyLysThrValSerValThrAla 737

2641 AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAA 2690
           ::::::   |||:::   :::|||::::::::::   |||   ::
 738 ThrThrAspSerLeuThrValLysGlyGlyAlaLysIle...AsnAlaTh 753

2691 CCTTGACAACGCCACCATTACACTCAAT..................... 2718
     :      :::|||||||:::||||   :::
 753 rGluGlyThrAlaThrLeuThrAlaSerSerGlyLysLeuThrThrGluA 770

2719 ..............TCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACC 2754
                  |||:::   :::       :::|||::::::|||:::
 770 laAsnSerAlaIleSerGlyAlaAsnGlyValThrAlaSerSerGlnSer 786

2755 GGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2804
     |||   :::|||   |||         ::::::
 787 GlyAspIleSerGlyThrIleSerGlyLysThr.................. 797

2805 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2854
                  :::||||||   :::      ::::::||||||||||:
 798 .............ValSerValThrAlaSerSerGlySerLeuThrValG 810

2855 ACGGC......AAATTGAAC...NGTCAAGGAACATTCCGCTTTATGTCG 2895
     ::|||      |||:::|||   :::::::|||:::      :::   :::
 810 lyGlyAspAlaLysIleAsnAlaThrGluGlyAlaAlaThrLeuThrAla 826

2896 GAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGA 2945
                       :::   |||   :::      ::::::::::||
 827 ThrLysGlyThrLeuThrThrValLysGlySerAsnIleAspAlaAsnGl 843

2946 AGGNACTTACACCTTG...GCGGTCAACAATACCGGCAACGAACCCGTAA 2992
     |||||||   ::::::   |||   :::   |||   |||      :::|
 843 uGlyThrLeuValIleAsnAlaGlnAspAlaThrLeuAsnGlyAspAlaS 860

2993 GCCTCGATCAATTGACGGTA............GTGGAAGGGAAAGACAAC 3030
     ||   |||:::      |||            :::::::|||:::   |||
 860 erGlyAspArgThrGluValAsnAlaValAsnAlaSerGlySerGlyAsn 876

3031 ......AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACA 3074
                    |||::::::::::|||:::|||   :::   :::
 877 ValThrAlaLysThrSerSerSerValAsnIleThrGlyAspLeuSerTh 893

3075 CGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGT 3124
      :::::::::                 ::::::|||   |||:::|||:::
 893 rIleAsnGly............LeuAsnIleIleSerLysAsnGlyLys. 905

3125 TCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGC 3174
               |||   |||                           |||
 906 ...........AsnThrVal....................ValLeuLys 911

3175 AAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3224
        ||||||         :::      :::   ::::::   |||   ::
 912 GlyAlaGluIleAspValLysTyrIleGlnProGlyValAlaSerAlaAs 928

3225 CGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTG 3274
     :   :::|||   |||   |||   |||   ||||||:::::::::::::
 928 nGluValIleGluAlaLysArg...AlaLeuGluLysValLysAspLeuS 944

3275 CCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3324
     :::::      |||   :::   ::::::   :::|||:::      :::
 944 erAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal 959

3325 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAA 3374
                  :::   ::::::   ::::::      ::::::
 960 .............ArgPheIleGluProAsnAsnThrIleThrValAsnTh 972

3375 ACAGCGCGAAGCGGAAACCCGGCCGGNTACC 3405
      |||:::|||      |||||||||:::::
 972 rGlnAsnGluPheThrThrArgProSerSer 982
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837

seq_documentation_block:

ID AAB01837 standard; Protein; 1073 AA.
XX
AC AAB01837;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain LCDC2 mature HMW2A protein, SEQ ID NO:49.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain LCDC2.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52186.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 23A-P; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
```
```

```
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1073 AA;


alignment_scores:
            Quality:  251.00              Length:   1090
              Ratio:  0.485                Gaps:     54
 Percent Similarity:  47.431   Percent Identity: 20.642

alignment_block:
 US-09-303-518D-651 x AAB01837   ..

 Align seg 1/1  to: AAB01837  from: 1  to: 1073

     772 GGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATG 821
         ||||||||||||      :::   :::|||   |||   :::|||:::
      74 GlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAsnArg.. 88

     822 GCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAA 871
               :::      :::|||   :::         ::::::      :
      89 .......SerAspGlyIleGlnIleAsnLysAspIleThrSerThrGlyG 103

     872 ACGGTTTCCAGCTGATACGCAAAGATTGG......TTCTACGATGACATT 915
         ::::::::   :::         ||||||      ::::::   :::|||
     103 lySerLeuThrIleAsnSerAspAspTrpValAspIleHisGlyAsnIle 119

     916 TACAGAGGCGAT.....................ACACATACCGTCTNTTT 944
               |||:::                     :::   :::|||:::||
     120 ThrLeuGlyGluGlyPheLeuAsnIleThrSerSerAspSerValAlaPh 136

     945 TGAA......CCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACA 988
         ||||            ::::::|||:::   |||   :::::::::
     136 eGluGlyGlyAsnGlyAsnLysGlyArgSerSerAlaSerAlaGlnIleI 153

     989 ACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAG 1038
           :::   ||||||:::|||   |||:::|||:::
     153 leAlaGlnGlyThrIleThrLeuThrGlyGluAsn.............. 164

    1039 CTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAACTGA 1088
                :::|||   ||||||   :::   |||||||||   |||
     165 .........LysThrPheArgLeuAsnAsnValSerLeuAsnGlyThr.. 177

    1089 TAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGGT 1138

     177 ..................................................... 177

    1139 TAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTC 1188
                                            |||||||||
     178 .................................GlyAsnGlyLeuSer 182

    1189 ATCTTATCAAACAACATCAACCAAGGC.........GCGGGCGGTTTGTA 1229
         |||:::|||:::      |||   :::          |||   :::
     183 IleIleSerThrAlaSerAsnLeuSerHisArgLeuAspGlyGluIleAs 199

    1230 TTTTGAAGGTGATTTTACGGTC.........TCGCCTGAAAACAACGAAA 1270
           :::|||:::   |||:::         :::   ::::||   |||
     199 nValSerGlyAsnValThrIleAsnGlnThrThrGlnGlnAsnIleGluT 216

    1271 CGTGGCAAGGC..................................... 1281
           |||::::::
     216 yrTrpLysAlaSerSerAspSerTyrTrpAsnValThrSerPheAsnLeu 232

    1281 ..................................................... 1281

     233 ArgGluAspSerLysPheThrPheIleLysTyrValAsnSerAlaArgAs 249

    1282 ........................GCGGGCGTT................. 1290
                                 |||||||||
     249 nGlyAspValArgGlyArgSerPheAlaGlyValIlePheAsnAlaLysG 266

    1291 .................CATATCAGTGAAGACAGTACCGTTACTTGGAAA 1323
                         ::::::::::::   |||||||||   :::|||
     266 lyLeuThrThrSerPheAsnValLysLysGlySerThrValAspPheLys 282

    1324 GTA......AACGGCGTGGCAAACGACCGC..................... 1347
         :::      |||:::      |||::::::
     283 LeuLysProAsnSerGlyTyrAsnSerGlnLysArgIleProIleGlnPh 299

    1348 ..........CTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCA 1387
                   :::|||      |||   |||   ::::::::::::::
     299 eGlnSerAsnIleSerValSerGlyGlyGlyArgValAsnIleAsnThrL 316

    1388 AAGGGGAAAACCAAGGC.......................TCGATCAGC 1413
         ::::::     |||                          ||||||:::
     316 euAlaAsnLeuThrGlyGlyGlyValGluIleArgSerSerSerIleAsn 332

    1414 GTGGGCGACGGT...ACAGTCATTTTGGATCAGCAGGCAGACGATAAAGG 1460
         |||:::||||||   |||:::   :::      ||||||   |||:::::
     333 ValSerAspGlySerThrLeuSerMetThrAlaGlnAlaArgAspArgAs 349

    1461 CAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGG 1510
         :      :::      :::::::::   :::   ::::::::::::::
     349 nAlaPheGluIleThrLysAspLeuValIleAsnAlaSerAsnSerAsnL 366

    1511 TGCAACTG.........................................AAT 1521
         ::::::::                                        |||
     366 euSerIleIleGlnGlnAsnAspGlyPheAspAsnAsnGlnLysAlaAsn 382

    1522 GCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGG 1571
         |||   |||            ::::||            ::::::||||
     383 AlaIleAsn............SerLysTyrAsnValThrIleGlnGlyGl 395

    1572 ACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATA 1621
         |::::::   |||:::|||                          ||||||:
     395 yAsnValThrLeuGlyGly.......................GlnAsnS 404

    1622 CCGATGAAGGGGCGATGATTGNCNATCATAATGCCACAACAACATCCACC 1671
         ::                          ::::::|||   |||::::::
     404 er.........................SerSerThrIleThrGlySer 411

    1672 GTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATAT 1721
         |||:::|||   :::|||   ::::::|||
     412 ValAsnIleGlyAlaAsnAlaAsnValThrLeuGln.............. 423

    1722 CAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCG 1771
                                  |||:::||||||      ::::
     424 ............................AlaHisAsnGly......AsnA 429

    1772 AGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCC 1821
         ::::::::         |||   |||:::::::::::::
     429 spArgAsnLysLysLeuThrPheGlyAsnValSerVal............ 441

    1822 GCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGG 1871
               |||   :::      |||:::::::::::::::|||:::|||::
     442 ......GluGlyGluLeuArgLeuValGlyAlaSerAlaAsnIleAsnAs 456

    1872 CAACATCACG..........................CAAACAAACGGCA 1894
         :|||::::::                          :::||||||   :
     456 nAsnLeuSerValLysSerGlyAlaLysPheLysAlaGluThrAsnAspA 473

    1895 AACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAAT......CAT 1938
```

```
    ::|||   ::::::|||      |||   ::::::   :::
 473 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrSerIleIleAsp 489

1939 TTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGT 1988
     :::   :::|||   :::|||:::              |||:::|||::
 490 ValLysLysGlyAlaAlaLysLeu.............GlyAsnIleTh 501

1989 GTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTCCATA 2038
     :  |||::::::   :::      |||   :::|||:::|||
 501 rAsnAspGlyAsn...LeuAsnIleThrThrAsnAlaLysAsn..... 514

2039 TTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGC 2088
         |||   ::::::||||||:::   |||:::::::::   :::|||
 515 .....GlyGlnLysSerValIleAsnGlyAsnIleThrAsnAsnLysGly 529

2089 GATTGNCATTTGAGCAATCACGCCCAA...........GCAGTTTTTGG 2126
           :::::::::|||:::::::::             :::   ||
 530 AlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleGlnIleGlyGl 546

2127 TGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNG 2176
     |         :::::::::   ::::::         ||||||      :
 546 yAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysIleA 563

2177 GTCTGACAAATTGTGTCGAANAAANCATT...ACCGAC........... 2211
     :::::|||:::   :::|||            ||||||
 563 snIleThrLysArgIleGluIleLysAlaGlyThrAspGlnGlyAsnSer 579

2212 GATAAAGTGATTGCTTCA.............................TT 2231
     |||:::   :::||||||                              ||
 580 AspSerGlyValAlaSerAsnAlaAsnLeuThrIleLysThrLysGluLe 596

2232 GACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTT 2281
     |      |||:::      :::       :::      :::
 596 uLysLeuThrGluAsnLeuAsnIleSerGlyPheAspLysAlaGluIleV 613

2282 NAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCAAATGGC 2331
       :::      :::         :::   ||||||   ::::::::::::
 613 alAlaLysGluAsnAsnAsnLeuIleIleGlyAsnAsnAsnGlyAspAsn 629

2332 GATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAG 2381
        :::      ||||||:::   |||   ::::::::::::::::||
 630 AlaAsnAlaLysThrValThrPheAsnAsnValLysAspSerLysIleSe 646

2382 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAA......GCCACATTAA 2425
     |      |||:::   ::::::|||:::|||:::         |||  :
 646 rAlaAsnGlyHis...AsnValThrLeuAsnSerLysValGluThrSerA 662

2426 ACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTAATCTA....... 2466
     ::||||||:::::::::::::||||||::::::   |||
 662 spGlyAsnSerAsnThrGluGlyAsnSerAspAsnAsnAlaGlyLeuThr 678

2467 ......AGCAACAACGCCGCACAAAACGGC..................... 2490
           ::::::|||::::::   |||:::
 679 IleAspAlaLysAsnValThrValAsnAsnAspIleThrSerHisLysTh 695

2491 ....AGTCTGACGCTTTCCGAC......AACGCTAAGGCAAACGTAAGCC 2530
        :::::::|||   |||:::      ::::::|||||||::::::::
 695 rValAsnIleThrAlaSerGluArgIleAspThrLysAlaAspThrThrI 712

2531 ATTCCGCACTCAACGGCAATGTCTCCCTA...................... 2559
       :::|||   :::|||||||||:::|||
 712 leAsnAlaThrThrGlyAsnValLysLeuThrAlaValThrSerAspIle 728

2560 ..............GCCGATAAGGCAGTATTCCATTTTGAAAACAGCCG 2594
                    ::::::::::::      ::::::   :::|||
 729 GlnGlyGlyIleLysSerAsnSerGlyAspValAsnIleThrThrSer.. 744

2595 CTTTACCGGACAACTCAGCGGC.........AGCAAGGANACAGCATTAC 2635
        ||||||:::::::::|||         ||||||:::   ::::::
 745 ....ThrGlySerIleAsnGlyLysIleGluSerLysSerGlySerValT 760

2636 ACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAAT 2685
       |||      :::|||   ||||||               :::||||||
```

```
 760 hrLeuThrAlaThrGluLysThrLeu.............ThrValGlyAsn 772

2686 TTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGA 2735
     ::::::   ::::::::::|||:::|||   |||   :::
 773 ValSerGlyAsnThrValThrValThrAlaAsnArgGlyAlaLeuThrTh 789

2736 TGCTGCAGGCGCGCAA...ACCGGCAGNGTGTCAGACACGCCGCGCCGCC 2782
          |||||||:::      :::|||:::   :::   |||        :
 789 rLeuAlaGlySerThrIleAsnGlyThrAsnGlyValThrThrSerSerG 806

2783 GTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT...TCGGTAGAA 2829
     ::|||   :::          :::||||||      |||   ||||||
 806 lnSerGlyGluIleGlyGlyGluValThrGlyLysThrValSerValThr 822

2830 TCCCGTTTCAACACGCTGACGGTAAACGGC......AAATTGAAC...NG 2870
     :::      ::::::||||||||||:::|||     |||:::|||   ::
 823 AlaThrAlaGlySerLeuThrValLysGlyGlyAlaLysIleAsnAlaTh 839

2871 TCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACA 2920
     :::::||||||      :::   ::::::                |||   |
 839 rGluGlyThrAlaThrLeuThrAlaSer..............SerGlyL 851

2921 AATTGAAGCTGGCGGAAAGTTCCGAAGGNACT.................. 2952
     |||||            ||||||:::   |||
 851 ysLeuThrThrGluAlaSerSerAsnIleThrSerAlaLysGlyGlnVal 867

2953 .............................................TACAC 2957
                                                    ::
 868 AspLeuSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAs 884

2958 CTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGA 3007
     ::::::::::|||:::||||||:::                    ||||
 884 nValThrLeuAsnThrThrGlyThr.....................LeuT 894

3008 CGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTC 3057
     ||:::||||||  |||:::::::   |||            |||||||:::
 894 hrThrValGlu.GlySerSerIleAsnAlaAsnGluGlyThrLeuValIl 910

3058 ACCCTGC.....................AAAACGAACACGTCGATGCCGG 3086
        :::                      |||::::::   :::
 910 eAsnAlaAsnAspAlaLysLeuAspGlyLysAlaSerGlyAsnArgThrG 927

3087 CGCGTGGCGTTACCAACTCATCC.................GCAAAGACG 3118
        ::::::||||||:::|||                  |||||||||
 927 luValAsnAlaThrAsnAlaSerGlySerGlySerValThrAlaLysThr 943

3119 GCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAA 3168
     :::|||||||:::   |||      :::   |||                ||
 944 SerSerSerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAs 960

3169 CTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3218
     |         ::::::   :::|||      :::         ::::::
 960 nIleIleSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluI 977

3219 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAA 3268
          ::::::   ::::::|||||||:::  |||:::      :::|||:
 977 leGluValLysTyrIleGlnProGlyVal..AlaSer......ValGluG 991

3269 GCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATG 3318
     ::|||   |||        |||          |||:::|||      :::
 991 luValIleGluAlaLysArgValLeu......GluLysValLysAspLeu 1005

3319 CAGGCGGAGGAAGAGAAA............................AA 3338
     :::   ||||||::::::                            ::
1006 SerAspGluGluArgGluThrLeuAlaLysLeuGlyValSerAlaValAr 1022

3339 ACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGG 3388
     :   ::::::   ::::::      ::::::      |||:::|||
1022 gPheIleGluProAsnAsnThrIleThrValAsnThrGlnAsnGluPheT 1039

3389 AAACCCGGCCGGNTACC 3405
       |||||||||::::::
1039 hrThrArgProSerSer 1044
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836

seq_documentation_block:
ID AAB01836 standard; Protein; 1079 AA.
XX
AC AAB01836;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain LCDC2 HMW2A protein, SEQ ID NO:47.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain LCDC2.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52185.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 23A-P; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 1079 AA;

alignment_scores:

Quality: 251.00 Length: 1090
Ratio: 0.485 Gaps: 54
Percent Similarity: 47.431 Percent Identity: 20.642

alignment_block:
US-09-303-518D-651 x AAB01836 ..

Align seg 1/1 to: AAB01836 from: 1 to: 1079

```
 772 GGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATG 821
     ||||||||||||      :::   :::|||   |||   :::|||:::
  80 GlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAsnArg.. 94

 822 GCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAA 871
           :::      :::|||   :::          ::::::      :
  95 .......SerAspGlyIleGlnIleAsnLysAspIleThrSerThrGlyG 109

 872 ACGGTTTCCAGCTGATACGCAAAGATTGG......TTCTACGATGACATT 915
     ::::::::   :::         ||||||      ::::::   :::|||
 109 lySerLeuThrIleAsnSerAspAspTrpValAspIleHisGlyAsnIle 125

 916 TACAGAGGCGAT.....................ACACATACCGTCTNTTT 944
           |||:::                     :::   :::|||:::||
 126 ThrLeuGlyGluGlyPheLeuAsnIleThrSerSerAspSerValAlaPh 142

 945 TGAA......CCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACA 988
     ||||          ::::::|||:::   |||   :::::::::
 142 eGluGlyGlyAsnGlyAsnLysGlyArgSerSerAlaSerAlaGlnIleI 159

 989 ACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAAG 1038
       :::   ||||||:::|||   |||:::|||:::
 159 leAlaGlnGlyThrIleThrLeuThrGlyGluAsn..............  170

1039 CTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAACTGA 1088
           :::|||   ||||||   :::   |||||||||   |||
 171 .........LysThrPheArgLeuAsnAsnValSerLeuAsnGlyThr.. 183

1089 TAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGGT 1138

 183 ..................................................  183

1139 TAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTC 1188
                                         |||||||||
 184 .................................GlyAsnGlyLeuSer 188

1189 ATCTTATCAAACAACATCAACCAAGGC.........GCGGGCGGTTTGTA 1229
     |||:::|||:::      |||   :::          |||   :::
 189 IleIleSerThrAlaSerAsnLeuSerHisArgLeuAspGlyGluIleAs 205

1230 TTTTGAAGGTGATTTTACGGTC.........TCGCCTGAAAACAACGAAA 1270
       :::|||:::   |||:::         :::   :::|||   |||
 205 nValSerGlyAsnValThrIleAsnGlnThrThrGlnGlnAsnIleGluT 222

1271 CGTGGCAAGGC..................................... 1281
       |||::::::
 222 yrTrpLysAlaSerSerAspSerTyrTrpAsnValThrSerPheAsnLeu 238

1281 ..................................................  1281

 239 ArgGluAspSerLysPheThrPheIleLysTyrValAsnSerAlaArgAs 255

1282 .........................GCGGGCGTT.................  1290
                               |||||||||
 255 nGlyAspValArgGlyArgSerPheAlaGlyValIlePheAsnAlaLysG 272

1291 ................CATATCAGTGAAGACAGTACCGTTACTTGGAAA 1323
                     ::::::::::::   |||||||||   :::|||
 272 lyLeuThrThrSerPheAsnValLysLysGlySerThrValAspPheLys 288

1324 GTA......AACGGCGTGGCAAACGACCGC.................... 1347
     :::      |||:::      |||::::::
 289 LeuLysProAsnSerGlyTyrAsnSerGlnLysArgIleProIleGlnPh 305
```



```
 845 rGluGlyThrAlaThrLeuThrAlaSer.............SerGlyL 857

2921 AATTGAAGCTGGCGGAAAGTTCCGAAGGNACT..... 2952
     |||||          |||||||:::   |||
 857 ysLeuThrThrGluAlaSerSerAsnIleThrSerAlaLysGlyGlnVal 873

2953 ........................................TACAC 2957
                                              ::
 874 AspLeuSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAs 890

2958 CTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGA 3007
     ::::::::::|||:::||||||:::                  ||||
 890 nValThrLeuAsnThrThrGlyThr.................LeuT 900

3008 CGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTC 3057
     ||:::|||||| |||:::::   |||           ||||||.:::
 900 hrThrValGlu.GlySerSerIleAsnAlaAsnGluGlyThrLeuValIl 916

3058 ACCCTGC.................AAAACGAACACGTCGATGCCGG 3086
         :::                 |||::::::   :::
 916 eAsnAlaAsnAspAlaLysLeuAspGlyLysAlaSerGlyAsnArgThrG 933

3087 CGCGTGGCGTTACCAACTCATCC.................GCAAAGACG 3118
         ::::::||||||:::|||                 |||||||||
 933 luValAsnAlaThrAsnAlaSerGlySerGlySerValThrAlaLysThr 949

3119 GCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAA 3168
     :::||||||:::   |||     :::   |||               ||
 950 SerSerSerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAs 966

3169 CTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3218
     |         ::::::   :::|||      :::         ::::::
 966 nIleIleSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluI 983

3219 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAA 3268
         ::::::   ::::::|||||||:::  |||:::      :::|||:
 983 leGluValLysTyrIleGlnProGlyVal..AlaSer......ValGluG 997

3269 GCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATG 3318
     ::|||   |||     |||           |||:::|||     :::
 997 luValIleGluAlaLysArgValLeu......GluLysValLysAspLeu 1011

3319 CAGGCGGAGGAAGAGAAA.............................AA 3338
     :::   ||||||::::::                            ::
1012 SerAspGluGluArgGluThrLeuAlaLysLeuGlyValSerAlaValAr 1028

3339 ACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGG 3388
     :   ::::::   ::::::      ::::::      |||:::|||
1028 gPheIleGluProAsnAsnThrIleThrValAsnThrGlnAsnGluPheT 1045

3389 AAACCCGGCCGGNTACC 3405
       |||||||||::::::
1045 hrThrArgProSerSer 1050
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731

seq_documentation_block:
ID   AAR41731 standard; Protein; 1338 AA.
XX
AC   AAR41731;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 3 (HMW3).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49510.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 5; Figure 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1338 AA;


alignment_scores:
           Quality:  251.00            Length:   1241
             Ratio:  0.418               Gaps:     69
 Percent Similarity:  48.429   Percent Identity: 21.515

alignment_block:
 US-09-303-518D-651 x AAR41731   ..

 Align seg 1/1  to: AAR41731  from: 1  to: 1338

  127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
      |||:::||||||:::|||                          :::::
   89 GlyGluThrTyrLeuGly.........................GlyAs 96

  177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACA 226
      :::::::|||:::       |||   :::  |||::::::   :::|
   96 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 110

  227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
      ||                        ::::::   |||:::      |||:::
  110 ys..................ThrThrLeuGluLysGlySerThrIleAsn 120

  277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||      :::::::::|||   |||   :::   ||||||
  121 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 137

  327 TATTGTG.........................AGCGTGGCACATAACG 349
       ::::::                        ::::::|||   :::|
  137 aLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrG 154

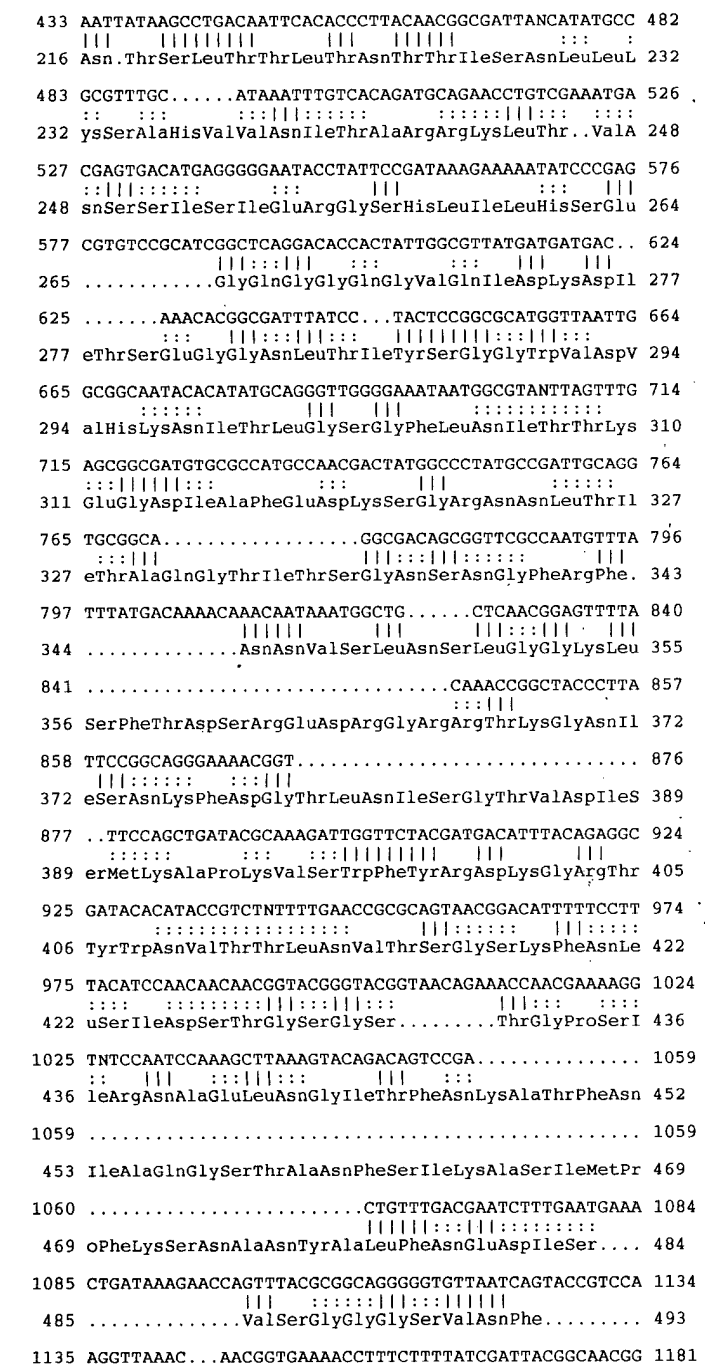
  350 GCGGCTAT..................................... 357
      |||||:::
  154 lyGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAla 170

  357 ..... 357

  171 IleValLysThrLysGluTrpLeuLeuAspProGluAsnValThrIleGl 187

  358 ............AACAACGTTGATTTTGGTGCGGAAGGAAGNAATCCCG 394
                  ::::::|||::::::||||||              |
  187 uAlaProSerAlaSerArgValGluLeuGlyAla..............A 199

  395 ATCAGCACCGTTTTTCTTACCAAATTGTG.....AAAAGAAAT 432
      ||::::::      |||   :::::::::           |||:::|||
  199 spArgAsnSerHisSerAlaGluValIleLysValThrLeuLysLysAsn 215
```


433 AATTATAAGCCTGACAATTCACACCCTTACAACGGCGATTANCATATGCC 482
216 Asn.ThrSerLeuThrThrLeuThrAsnThrThrIleSerAsnLeuLeuL 232
483 GCGTTTGC......ATAAATTTGTCACAGATGCAGAACCTGTCGAAATGA 526
232 ysSerAlaHisValValAsnIleThrAlaArgArgLysLeuThr..ValA 248
527 CGAGTGACATGAGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAG 576
248 snSerSerIleSerIleGluArgGlySerHisLeuIleLeuHisSerGlu 264
577 CGTGTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGAC.. 624
265 GlyGlnGlyGlyGlnGlyValGlnIleAspLysAspIl 277
625 AAACACGGCGATTTATCC...TACTCCGGCGCATGGTTAATTG 664
277 eThrSerGluGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAspV 294
665 GCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTG 714
294 alHisLysAsnIleThrLeuGlySerGlyPheLeuAsnIleThrThrLys 310
715 AGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGG 764
311 GluGlyAspIleAlaPheGluAspLysSerGlyArgAsnAsnLeuThrIl 327
765 TGCGGCA.................GGCGACAGCGGTTCGCCAATGTTTA 796
327 eThrAlaGlnGlyThrIleThrSerGlyAsnSerAsnGlyPheArgPhe. 343
797 TTTATGACAAAACAAACAATAAATGGCTG......CTCAACGGAGTTTTA 840
344 AsnAsnValSerLeuAsnSerLeuGlyGlyLysLeu 355
841 CAAACCGGCTACCCTTA 857
356 SerPheThrAspSerArgGluAspArgGlyArgArgThrLysGlyAsnIl 372
858 TTCCGGCAGGGAAAACGGT.............................. 876
372 eSerAsnLysPheAspGlyThrLeuAsnIleSerGlyThrValAspIleS 389
877 ..TTCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGC 924
389 erMetLysAlaProLysValSerTrpPheTyrArgAspLysGlyArgThr 405
925 GATACACATACCGTCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTT 974
406 TyrTrpAsnValThrThrLeuAsnValThrSerGlySerLysPheAsnLe 422
975 TACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAACGAAAAGG 1024
422 uSerIleAspSerThrGlySerGlySer.........ThrGlyProSerI 436
1025 TNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGA.............. 1059
436 leArgAsnAlaGluLeuAsnGlyIleThrPheAsnLysAlaThrPheAsn 452
1059 ... 1059
453 IleAlaGlnGlySerThrAlaAsnPheSerIleLysAlaSerIleMetPr 469
1060 CTGTTTGACGAATCTTTGAATGAAA 1084
469 oPheLysSerAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSer.... 484
1085 CTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCA 1134
485 ValSerGlyGlyGlySerValAsnPhe......... 493
1135 AGGTTAAAC...AACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGG 1181
494 LysLeuAsnAlaSerSerSerAsnIleGln............ThrProGl 506
1182 CAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGC.........G 1222
506 yValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerThrLeuA 523
1223 GTTTGTATTTTGAAGGTGAT............TTTACGGTCTCGCCTGAA 1260
523 snLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGluAsnAsp 539
1261 AACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTAC 1310
540 LeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGln........ 553
1311 CGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCG 1360
554 ValGluGlyThrAsp............SerArgValA 562
1361 GCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATC 1410
562 snLysGly.........ValAlaAlaLys............LysAsnIle 571
1411 AGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGG 1460
572 ThrPheLysGlyGlyAsnIleThrPhe.....................Gl 581
1461 CAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGG 1510
581 ySerGlnLysAlaThrThrGluIle..............LysGlyAsnV 593
1511 TGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC 1560
593 alThrIleAsnLysAsnThrAsnAlaThr..................... 602
1561 TTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCG 1610
603 LeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuAsnIle..... 617
1611 TATTCAAAATACCGATGAAGGGGCGATGATTGNCNATCATAATGCCACAA 1660
618 AlaGlyAsnValIleAsnAsnGlyAsnLeuThrT 629
1661 CAACA...TCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCG 1707
629 hrAlaGlySerIleIleAsnIleAlaGlyAsn......LeuThrValSer 643
1708 AGTGGTAAGAATATCAATAGACTT...AATTACAGCAAAGAAATTGCCTA 1754
644 LysGlyAlaAsnLeuGlnAlaIleThrAsnTyrThrPheAsnValAla.. 659
1755 CAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCA 1804
660 GlySerPheAsp.........AsnAsnGlyAlaSerA 669
1805 ACCTT...GTTTACCAGCCCGCCGCAGAA............GACCGCACC 1839
669 snIleSerIleAlaArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 685
1840 CNGCTGCTTTCCGGCGGAACAAAT.......................TT 1865
686 SerSerLeuAsnIleThrThrAsnSerAspThrThrTyrArgThrIleIl 702
1866 AAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGAC 1915
702 eLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLysL 719
1916 CGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAA 1965
719 ysSer...AspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGlu 734
1966 GGT.........ATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGAT 2006

```
 735 GlyAsnLeuThrIleSerSerAspLysValAsnIleThrAsnGln..... 749

2007 CNACCGCACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCG. 2055
        |||:::||||||           ::::::||||||::::::
 750 ....IleThrIleLysAla.........GlyValGluGlyGlyArgSerA 762

2056 ............................GTGATTTCCCGCAATGTTGCC 2076
                                 :::|||:::   :::
 762 spSerSerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeu 778

2077 AAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGG 2126
     |||:::   ||||||   ::::::|||:::      :::|||   :::
 779 LysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleTh 795

2127 TGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNG 2176
      :::         ::::::   ||||||   :::   |||
 795 rAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGly....... 809

2177 GTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGCT 2226
            |||:::              :::|||   |||||||||
 810 .....GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAsp 824

2227 TCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGN 2276
     |||      :::||||||      :::         ::::::     ::
 825 SerLysIleSerThrAspGlyHisAsn.........ValThrLeuAsnSe 838

2277 TNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCAA 2326
     :           ::::::   ::::::   :::||||||   :::::::
 838 rGluValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerT 855

2327 ATGGCGATACA............CGTTATACAGTCAGCCACAACGCCACC 2364
     ::|||   |||                |||||||:::::|||:::|||
 855 hrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnValThr 871

2365 CAAAAC..........................GGCAACCTT........ 2379
     ::::::                         |||||||:::
 872 SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrThrLy 888

2380 ...................AGCCTCGTGGGCAATGCCCAAGCAACATTTA 2410
                        :::   :::|||:::::::::::::|||   :
 888 sGluGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAlaG 905

2411 ATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTT 2460
     ::::::::|||:::::||||||   :::                 |||
 905 lnAsnGlyThrIleLysGlyAsnIleThr..............SerGln 916

2461 AATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAA 2510
     |||::::::   :::|||:::::|||      |||:::   ::::::||
 917 AsnValThrValThrAlaThrGluAsn......LeuValThrThrGluAs 931

2511 CGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAG 2560
     ||||      |||            |||   :::|||:::|||:::::::
 931 nAlaValIleAsn............AlaThrSerGlyThrValAsnIleS 944

2561 CCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTC 2610
     ::   |||                                |||||:::::
 944 erThrLys............................ThrGlyAspIle 950

2611 AGCGGC.........AGCAAGGANACAGCATTACACTTAAAAGACAGCGA 2651
     :::|||         |||   :::      :::::::::      |||
 951 LysGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGl 967

2652 ATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACG 2701
         ||||||           :::::::::|||:::::   :::::::
 967 yAsnThrLeu............LysValSerAsnIleThrGlyGlnAspV 980

2702 CCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCG... 2748
     ::|||:::|||   :::::::::           :::|||||||:::
 980 alThrValThrAlaAspAlaGlyAlaLeuThrThrThrAlaGlySerThr 996

2749 ................................CAAACCGGCAGNGT 2762
                                     :::|||||||   ::
 997 IleSerAlaThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIl 1013

2763 GTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2812
     ::::          :::      |||          |||   :::|||
1013 eAsnGlyLysValGluSerSerSerGlySerValThrLeuValAlaThrG 1030

2813 CGCCAACTTCGGTAGAATCCCGTTTC......AACACGCTGACGGTA... 2853
          |||   :::   :::::::::      ||||||:::|||:::
1030 lyAlaThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThr 1046

2854 ......AACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGA 2897
           :::|||||||||           |||            |||
1047 AlaAspSerGlyLysLeu............ThrSerThrValGlySerTh 1059

2898 ACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCC.... 2943
      :::   |||   :::|||              ::::::   ||||||
1059 rIleAsnGlyThrAsnSer............ValThrThrSerSerGlnS 1072

2944 ...........GAAGGNACTTACACCTTGGCGGTCAACAATACCGGCAAC 2982
                 |||||||||   :::         :::||||||   |||
1072 erGlyAspIleGluGlyThrIleSer.........GlyAsnThrValAsn 1085

2983 GAACCCGTAAGCCTCGATCAATTGACGGTA..............GTGGA 3017
           :::|||      :::|||||||:::               |||||
1086 ValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysValGl 1102

3018 AGGGAAAGACAACAAACCG......CTGTCCGAAAACCTTAATTTCACCC 3061
     |:::|||::::::              :::|||:::   :::::::|||
1102 uAlaLysAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrT 1119

3062 TGCAAAACGAACACGTCGATG...CCGGCGCGTGGCGTTACCAACTCATC 3108
       ||| |||:::::::|||:::      ::::::|||   |||:::   ::
1119 hrGln.ThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuTh 1135

3109 CGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCT 3158
     :||||||   :::|||   |||      |||::::::   |||   ::::
1135 rAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrL 1152

3159 TTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACA 3208
     ::   |||:::::::::                :::   :::   ::::::
1152 euAsnThrThrGlyThrLeuThrThrThrGlyAspSerLysIleAsnAla 1168

3209 ACGCGCAAAGCCTTGACG 3226
     |||      :::||||||
1169 ThrSerGlyThrLeuThr 1174
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728

seq_documentation_block:

```
ID   AAR41728 standard; Protein; 1477 AA.
XX
AC   AAR41728;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 2 (HMW2).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
```

```
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49509.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41724 (High molecular weight protein 2)
CC   although it does differ slightly.
XX
SQ   Sequence   1477 AA;
```

```
alignment_scores:
           Quality:  251.00            Length:   1450
             Ratio:  0.382             Gaps:       67
 Percent Similarity:  45.310   Percent Identity: 19.172

alignment_block:
 US-09-303-518D-651 x AAR41728   ..

 Align seg 1/1  to: AAR41728  from: 1  to: 1477

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
         :::|||          |||   :::   |||:::::::::::::   |||||
     158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

     213 TGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
         |:::      |||   |||::::::            ::::::|||||
     174 eSerAsnGluAsnIleLysAlaArgAsnPheThrPheGluGlnThrLysA 191

     262 ..........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                 |||   :::::::::   ::::::      |||:::::::
     191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

     295 AACGGCGTGGCGGCATTGGTGGGC............GATCAATATATTGT 332
         :::|||   :::   |||:::|||           ::::::   :::::
     208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

     333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT....... 375
         :||||||      |||||||||      ::::::      |||
     224 eSerVal......AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

     376 ........GCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAA 417
                 ::::::      ||||||           ::::::|||:::
     239 leThrIleSerAspIleIleAsnPro.........ThrIleThrTyrSer 252

     418 ATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGG 467
         |||:::                  |||:::|||:::            ||
     253 IleAla..............AlaProGluAsnGluAlaValAsnLeuGl 264

     468 CGAT.................................TANCATATGC 481
         ||||                              :::   :::
     264 yAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleA 281

     482 CGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGT 531
           :::      |||:::   :::|||:::           :::::::::
     281 rgAsnGlnGlyLysLeuSerAlaAspSer............ValSerLys 293

     532 GACATGAGGGGGAAT...ACCTATTCCGATAAAGAAAAATATCCCGAGCG 578
         |||      ||||||   ::::   |||   ||||||         |||
     294 AspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIl 310

     579 TGTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAAC 628
               :::   |||:::::::::           ::::::   |||
     310 eGlyGlyValIleSerAlaGlnAsn............GlnGlnAlaLysG 323

     629 ACGGCGATTTATCCTACTCCGGCGCATGGTTAATT.............. 663
           |||   |||      :::|||      :::
     323 lyGlyLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAla 339

     664 .......................GGCGGCAATACACATATG........ 681
                                 ||||||:::|||::::::
     340 ValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAs 356

     682 ....CAGGGTTGGGGAAATAATGGCGTA..................... 705
             :::|||   |||:::||||||:::
     356 pGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerL 373

     706 ................NTTAGTTTGAGCGGCGATGTGCGC......CAT 732
                         ::::::||||||      :::      :::
     373 euGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyArg 389

     733 GCCAACGACTATGGCCCTATGCCGATT.................GCAGG 764
         |||      :::|||   :::   :::                  |||
     390 AlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGl 406

     765 TGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACA 814
          ::::::||||||   :::                        ||||||
     406 nGlySerGlyAspIleAla.....................LysThr.... 414

     815 ATAAATGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGC 864
                    :::|||   ::::::|||:::      |||   :::
     415 ..............GlyGlyPheValGluThrSerGlyHisTyrLeuSer 426

     865 AGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGGTTCTACGAT..... 909
           :::::::::      :::      |||:::|||:::   |||
     427 IleAspSerAsnAlaIleValLysThrLysGluTrpLeuLeuAspProAs 443

     910 ......................GACATTTACAGAGGCGATACACATACCG 937
                                |||      |||:::::::|||
     443 pAspValThrIleGluAlaGluAspProLeuArgAsnAsnThrGlyIleA 460

     938 TCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAAC 987
                   |||   ::::::|||:::   |||      :::::::|||
     460 snAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLysAsn 476

     988 AACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAA 1037
         :::          |||   |||   ||||||      :::||||||
     477 SerGluLeuLysThr...ThrLeuThrAsnThrThrIleSerAsnTyrLe 492

    1038 GCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAACTG 1087
         ::::::   |||:::                  :::::::::   :::
     492 uLysAsnAlaTrpThrMet..............AsnIleThrAlaSerA 504

    1088 ATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGG 1137
           |||      |||   ::::::
     504 rgLysLeuThrValAsnSerSer........................... 511

    1138 TTAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACT 1187
         :::|||   |||:::|||      :::|||   ::::::::|||
     512 IleAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526

    1188 CATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAG 1237
                              |||   :::|||||||:::   ::::::|
     527 .....................GlnArgGlyGlyGlyValGlnIleAspG 536

    1238 GTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGC 1287
         |||||:::|||   :::      |||         ::::::|||:::
     536 lyAspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrp 552

    1288 GTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGC 1337
         |||   :::   ::::::   |||:::
     553 ValAspValHisLysAsnIleThrLeu....................... 561

    1338 AAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCA 1387

     561 ..................................................... 561
```

```
1388 AAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTG 1437
562 ........AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe 575

1438 GATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGG 1487
576 ...GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.. 590

1488 CTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCA 1537
591 .......ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys.... 603

1538 ACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAAC 1587
604 ...................AspPheArgAlaAsnAsnValSerLeuAsn 613

1588 GGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGAT 1637
614 GlyThrGlyLysGlyLeuAsnIleIleSerSerValAsnAsn........ 627

1638 GATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGA 1687
628 ....LeuThrHisAsn......LeuSerGlyThrIleAsnIleSerGlyA 641

1688 ATGAAAGTATTACACAACCGAGTGGTAAG..................... 1716
641 snIleThrIleAsnGlnThrThrArgLysAsnThrSerTyrTrpGlnThr 657

1717 .................AATATCAATAGACTTAATTACAGCAAAGAAAT 1748
658 SerHisAspSerHisTrpAsnValSerAlaLeuAsnLeu......GluTh 672

1749 TGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGC 1798
672 rGlyAlaAsnPheThrPheIleLysTyrIleSerSerAsnSerLysGly. 688

1799 GGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTT 1848
689 ..LeuThrThrGlnTyrArgSerSerAla..................... 697

1849 TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACT 1898
698 ......GlyValAsnPheAsnGly............ValAsnGlyAsnMe 708

1899 GTTTTTC..................................................... 1905
708 tSerPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProA 725

1906 .............AGCGGCAGACCGACACCGCACGCCTAC..... 1932
725 snGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsn 741

1933 ...........................................AATCA 1937
742 IleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsnHi 758

1938 TTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCG 1987
758 sSerGlyArgGly............AlaGluLeuLysMetSerGluIleA 771

1988 TGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTC... 2034
771 snIleSerAsn.....................GlyAlaAsnPheThr 779

2035 .........CATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGC 2075
780 LeuAsnSerHisValArgGlyAspAspAlaPhe.............. 790

2076 CAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTG 2125
791 .LysIleAsnLysAspLeuThrIleAsn..................... 799

2126 GTGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACN 2175

800 .............................AlaThrAsnSerAsnPhe... 805

2176 GGTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGC 2225

805 ..................................................... 805

2226 TTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACG 2275
806 .SerLeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaArgAsnA 822

2276 NTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCA 2325
822 laIleAsnSerThrTyrAsnIleSerIleLeuGlyGlyAsnValThrLeu 838

2326 AATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAA 2375
839 GlyGly.............................GlnAsnSerSe 844

2376 CCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCC....... 2418
844 rSerSerIleThrGlyAsn......IleThrIleGluLysAlaAlaAsnV 859

2419 ..ACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTAATCTA 2466
859 alThrLeuGluAlaAsnAsnAlaProAsnGlnGln.........AsnIle 872

2467 AGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAA 2516
873 ArgAspArgValIleLysLeuGlySerLeuLeuValAsnGlySerLeuSe 889

2517 GGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATA 2566
889 rLeuThrGlyGluAsnAlaAspIleLysGlyAsnLeuThrIleSerGluS 906

2567 AGGCAGTATTCCAT.........TTTGAAAACAGCCGCTTTACCGGACAA 2607
906 erAlaThrPheLysGlyLysThrArgAspThrLeuAsnIleThrGlyAsn 922

2608 CTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGAC 2657
923 PheThrAsnAsnGlyThrAlaGluIleAsnIleThrGlnGlyValValLy 939

2658 GCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCA 2707
939 sLeuGlyAsnValThrAsnAspGlyAspLeuAsnIle............. 951

2708 TTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGC 2757

951 ..................................................... 951

2758 AGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGT 2807
952 .........ThrThrHisAlaLysArgAsnGlnArgSerIleIleGly.. 964

2808 TACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACG 2857
965 ............................GlyAspIleIleAsnLysLysG 972

2858 GCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCG...GAACTCTTC 2904
972 lySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGly 988

2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
989 GlyAsnIleSerGlnLys.....................GluGlyAsnLe 998

2955 CACCTTGGCG.........GTCAACAATACCGGCAACGAACCCGTAAGCC 2995
998 uThrIleSerSerAspLysIleAsnIleThr.................. 1008

2996 TCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCC... 3042
```

```
1009 ..LysGlnIleThrIleLysLysGlyIleAspGlyGluAspSerSerSer 1024

3043 .........GAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGC 3083
                ::::|||   |||:::|||::::::
1025 AspAlaThrSerAsnAlaAsnLeuThrIleLys..................  1035

3084 CGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGC 3133

1035 ..................................................  1035

3134 ATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTC............  3171
            :::||||||   :::|||::::::   |||
1036 ........ThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGly 1049

3172 ...GGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3218
       :::|||||||||            :::   ||||||
1050 PheAsnLysAlaGluIle.........ThrAlaLysAsp...........  1059

3219 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAA 3268
                              |||||||||
1060 ......................GlyArgAspLeu.................  1063

3269 GCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATG 3318
                           :::   |||::::::|||   |||
1064 ...................ThrIleGlyAsnSerAsnAspGlyAsnSer 1073

3319 CAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTT 3368
       ||||||   :::         :::      |||||||||   ::
1074 GlyAlaGluAlaLysThrValThrPheAsnAsnValLysAspSerLysIl 1090

3369 GGCGAAACAGCGCGAAGCGGAAACC.........CGGCCGGNTACCACCG 3409
     ::::   :::   :::      |||         :::      |||::::
1090 eSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerS 1107

3410 CCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCC 3459
     ::         :::|||   ::::::   |||
1107 erSerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuThr 1123

3460 CAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGC...CGTTA 3506
                :::   :::   ::::::|||::::::|||   :::
1124 IleThrAlaLysAsnValGluValAsnLysAspIleThrSerLeuLysTh 1140

3507 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3556
      :::|||         ||||||      :::||| :::::::::   :::
1140 rValAsnIleThrAlaSerGluLysValThrThr.ThrAlaGlySerThr 1156

3557 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCN 3606
            ||||||      |||   :::         ::::::
1157 IleAsnAlaThrAsnGlyLysAlaSerIleThrThrLysThrGlyAspIl 1173

3607 GTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGC...AAGA 3653
       |||      :::::::::|||   ::::::::::::         :
1173 eSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAspL 1190

3654 TTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGA 3703
     ::::::::   :::::::::|||      :::   ::::::   ::::::
1190 euThrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsn 1206

3704 AAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3753
       |||||||||:::|||:::   ::::::         |||   :::
1207 ValThrSerAlaThrGlyThrIleGlyGly.........ThrIleSerGl 1220

3754 GAAAACANCTTCGACGACGGCATCGGCA..................... 3781
      :::|||   ::::::||||||:::|||
1220 yAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnG 1237

3782 ..............ACTCGGCACGGCTTGCCCACGGCGCCCGTTTTCGGGC 3817
                    :::   :::|||      |||         :::|||
1237 lyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrGly 1253

3818 AAT......ACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCG 3861
     |||      |||::::::||||||||||:::   :::|||::::::::::
1254 AsnThrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGl 1270

3862 GGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCG 3911
       :|||         |||      |||::::::|||   |||   |||
1270 nValAspLeuLeuAla......GlnAsnGlySerIleAlaGlySerIleA 1285

3912 CCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCG 3961
       |||:::   :::|||               |||      :::   |||
1285 snAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAla 1301

3962 GATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAA 4009
       |||      :::   ||||||:::      :::|||::::::|||
1302 GlySerAspIleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506

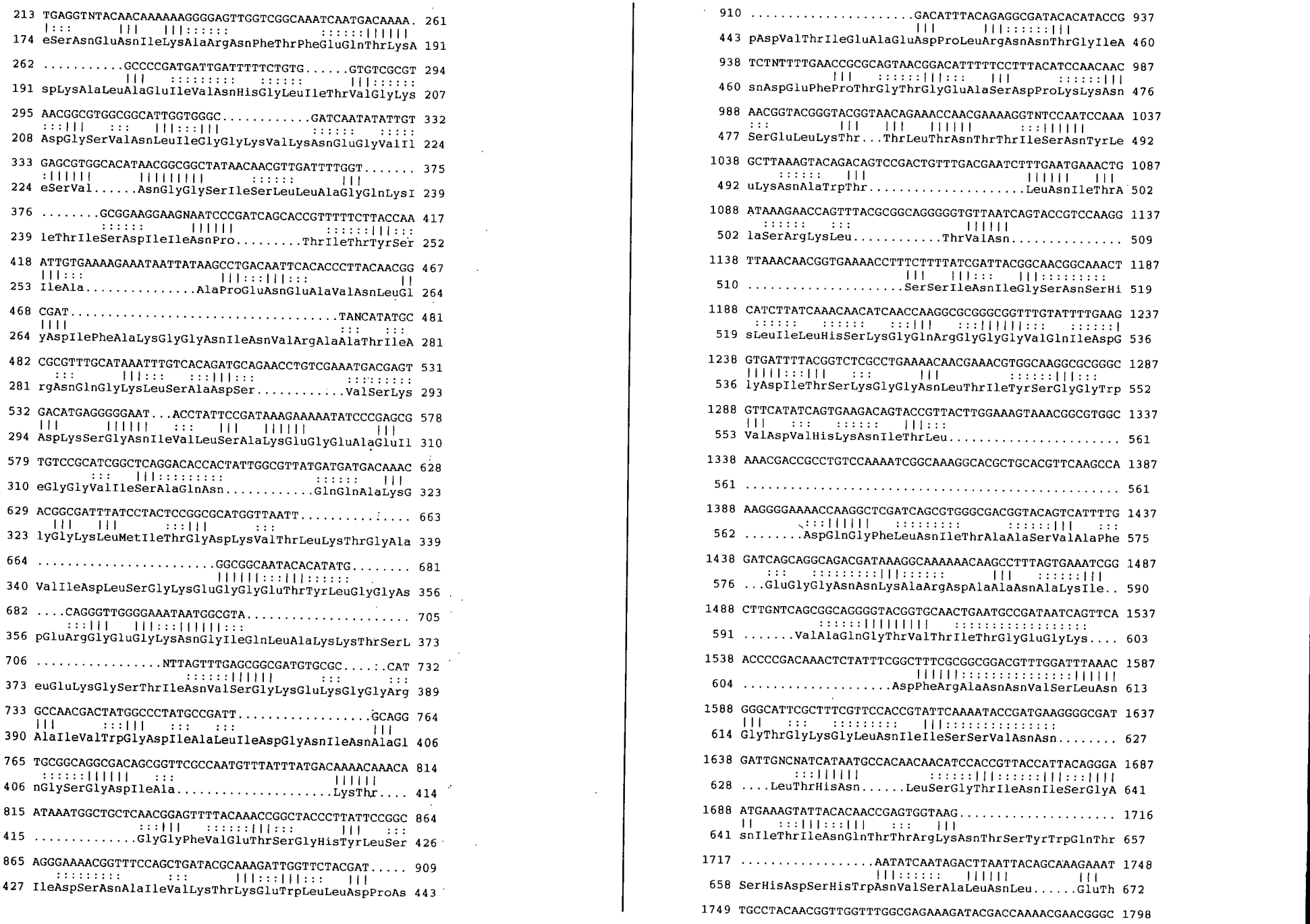
```
seq_documentation_block:
ID   AAR63506 standard; Protein; 1477 AA.
XX
AC   AAR63506;
XX
DT   25-JUN-1995  (first entry)
XX
DE   Haemophilus high molecular weight protein HMW2.
XX
KW   High molecular weight protein; HMW2; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib; ss.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72294.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 3; Page 36; 127pp; English.
XX
CC   The HMW2 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
SQ   Sequence   1477 AA;


alignment_scores:
            Quality:  251.00              Length:   1447
              Ratio:  0.380                 Gaps:     65
 Percent Similarity:  45.612   Percent Identity: 19.143

alignment_block:
 US-09-303-518D-651 x AAR63506   ..

 Align seg 1/1  to: AAR63506  from: 1  to: 1477

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
           :::|||         |||   :::   |||::::::::::::   |||||
     158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174
```





```
 3419 GCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAA 3468
      :::|||   ::::::   |||
 1110 lyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuThrIleThrAla 1126

 3469 CCTCAACCCCAACCGCAGCGCGACCTGATNAGC...CGTTATGCCAATAG 3515
         :::   :::   ::::::|||:::::::|||   :::   :::|||
 1127 LysAsnValGluValAsnLysAspIleThrSerLeuLysThrValAsnIl 1143

 3516 CGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGG 3565
             ||||||      :::||| :::::::::   :::
 1143 eThrAlaSerGluLysValThrThr.ThrAlaGlySerThrIleAsnAla 1159

 3566 ACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACA 3615
      ||||||     |||   :::         ::::::         |||
 1160 ThrAsnGlyLysAlaSerIleThrThrLysThrGlyAspIleSerGlyTh 1176

 3616 AGCNGCATCCGGNACACCAAACACTACCGTTCGC...AAGATTTCCGCGC 3662
         ::::::::|||   ::::::::::::         :::::::::
 1176 rIleSerGlyAsnThrValSerValSerAlaThrGlyAspLeuThrThrL 1193

 3663 CTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCG 3712
        ::::::::|||      :::   ::::::   ::::::   ||||||
 1193 ysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsnValThrSer 1209

 3713 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANC 3762
      |||:::|||:::   ::::::         |||   :::   :::|||
 1210 AlaThrGlyThrIleGlyGly.........ThrIleSerGlyAsnThrVa 1223

 3763 TTCGACGACGGCATCGGCA.............................. 3781
       ::::::||||||:::|||
 1223 lAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluI 1240

 3782 .....ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT..... 3820
          :::   :::|||      |||         :::||||||
 1240 leAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnThrLeu 1256

 3821 ACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGC 3870
      |||::::::||||||||||:::   :::|||:::::::::::::|||
 1257 ThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnValAspLe 1273

 3871 AGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCT 3920
         |||      |||::::::|||   |||   |||      |||:::
 1273 uLeuAla......GlnAsnGlySerIleAlaGlySerIleAsnAlaAlaA 1288

 3921 GCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCA 3970
        :::|||               |||      :::   |||   |||
 1288 snValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAsp 1304

 3971 TCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAA 4009
        :::   |||||||:::      :::|||:::::::|||
 1305 IleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848

seq_documentation_block:

```
ID   AAB01848 standard; Protein; 1477 AA.
XX
AC   AAB01848;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW2A protein, SEQ ID NO:71.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52197.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 29A-N; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1477 AA;
```

alignment_scores:

```
            Quality:  251.00              Length:   1450
              Ratio:  0.382                 Gaps:     67
 Percent Similarity:  45.310   Percent Identity: 19.172
```

alignment_block:

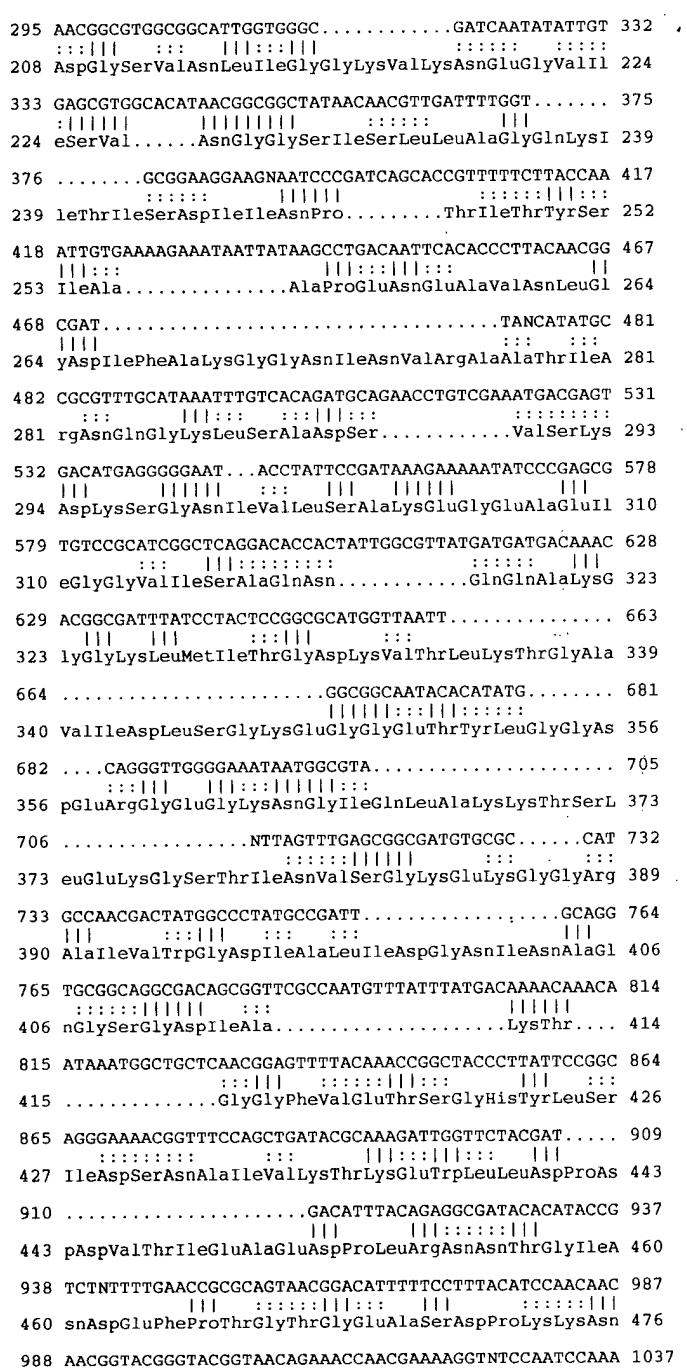
US-09-303-518D-651 x AAB01848 ..

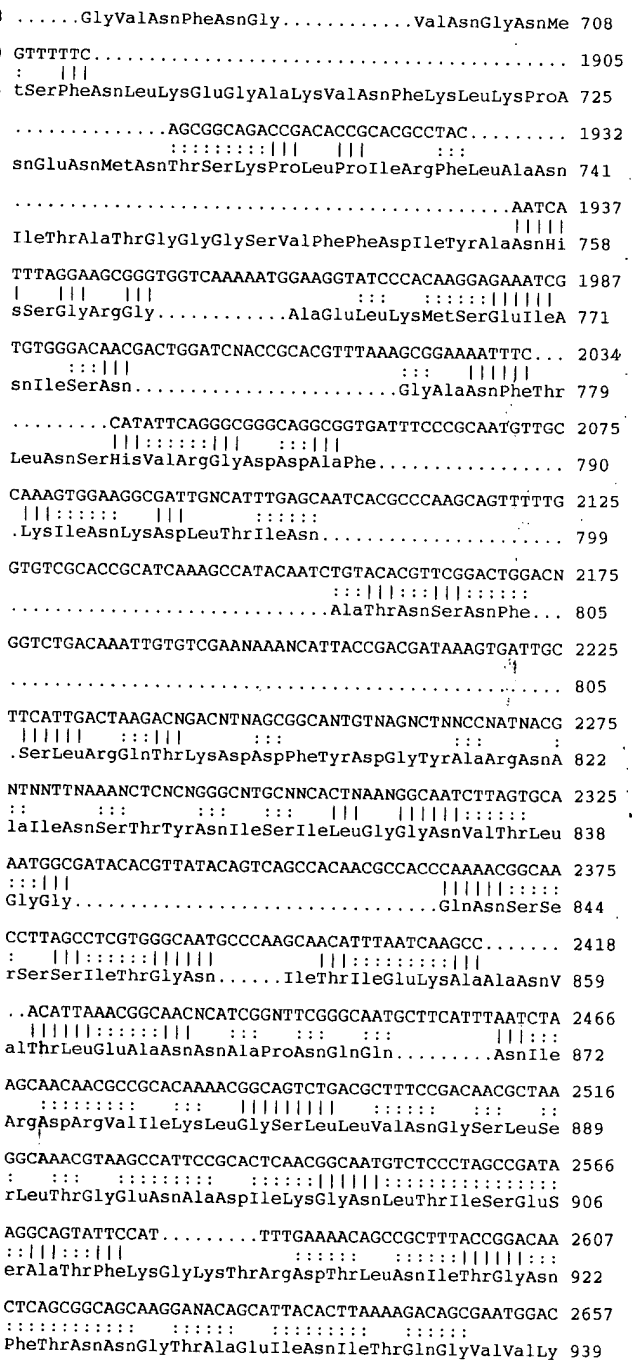
Align seg 1/1 to: AAB01848 from: 1 to: 1477

```
  163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
      :::|||         |||   :::   |||:::::::::::::   ||||| .
  158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

  213 TGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
      |:::       |||   |||::::::           ::::::||||||
  174 eSerAsnGluAsnIleLysAlaArgAsnPheThrPheGluGlnThrLysA 191

  262 ..........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                 |||   :::::::::   ::::::      |||::::::
  191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207
```


295 AACGGCGTGGCGGCATTGGTGGGC.....GATCAATATATTGT 332
208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224
333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT..... 375
224 eSerVal.....AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239
376 GCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAA 417
239 leThrIleSerAspIleIleAsnPro.....ThrIleThrTyrSer 252
418 ATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGG 467
253 IleAla.....AlaProGluAsnGluAlaValAsnLeuGl 264
468 CGAT.....TANCATATGC 481
264 yAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleA 281
482 CGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGT 531
281 rgAsnGlnGlyLysLeuSerAlaAspSer.....ValSerLys 293
532 GACATGAGGGGGAAT...ACCTATTCCGATAAAGAAAAATATCCCGAGCG 578
294 AspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIl 310
579 TGTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGATGATGACAAAC 628
310 eGlyGlyValIleSerAlaGlnAsn.....GlnGlnAlaLysG 323
629 ACGGCGATTTATCCTACTCCGGCGCATGGTTAATT..... 663
323 lyGlyLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAla 339
664 GGCGGCAATACACATATG..... 681
340 ValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAs 356
682 CAGGGTTGGGGAAATAATGGCGTA..... 705
356 pGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerL 373
706 NTTAGTTTGAGCGGCGATGTGCGC.....CAT 732
373 euGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyArg 389
733 GCCAACGACTATGGCCCTATGCCGATT.....GCAGG 764
390 AlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGl 406
765 TGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACA 814
406 nGlySerGlyAspIleAla.....LysThr.... 414
815 ATAAATGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGC 864
415 GlyGlyPheValGluThrSerGlyHisTyrLeuSer 426
865 AGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGGTTCTACGAT..... 909
427 IleAspSerAsnAlaIleValLysThrLysGluTrpLeuLeuAspProAs 443
910 GACATTTACAGAGGCGATACACATACCG 937
443 pAspValThrIleGluAlaGluAspProLeuArgAsnAsnThrGlyIleA 460
938 TCTNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAAC 987
460 snAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLysAsn 476
988 AACGGTACGGGTACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAA 1037
477 SerGluLeuLysThr...ThrLeuThrAsnThrThrIleSerAsnTyrLe 492
1038 GCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAACTG 1087
492 uLysAsnAlaTrpThrMet.....AsnIleThrAlaSerA 504
1088 ATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGG 1137
504 rgLysLeuThrValAsnSerSer..... 511
1138 TTAAACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACT 1187
512 IleAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1188 CATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAG 1237
527 GlnArgGlyGlyGlyValGlnIleAspG 536
1238 GTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGC 1287
536 lyAspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrp 552
1288 GTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGC 1337
553 ValAspValHisLysAsnIleThrLeu..... 561
1338 AAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCA 1387
561 561
1388 AAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTG 1437
562 AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe 575
1438 GATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGG 1487
576 ...GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.. 590
1488 CTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCA 1537
591 ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys.... 603
1538 ACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAAC 1587
604 AspPheArgAlaAsnAsnValSerLeuAsn 613
1588 GGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGAT 1637
614 GlyThrGlyLysGlyLeuAsnIleIleSerSerValAsnAsn..... 627
1638 GATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATTACAGGGA 1687
628 LeuThrHisAsn.....LeuSerGlyThrIleAsnIleSerGlyA 641
1688 ATGAAAGTATTACACAACCGAGTGGTAAG..... 1716
641 snIleThrIleAsnGlnThrThrArgLysAsnThrSerTyrTrpGlnThr 657
1717 AATATCAATAGACTTAATTACAGCAAAGAAAT 1748
658 SerHisAspSerHisTrpAsnValSerAlaLeuAsnLeu.....GluTh 672
1749 TGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGC 1798
672 rGlyAlaAsnPheThrPheIleLysTyrIleSerSerAsnSerLysGly. 688
1799 GGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTT 1848
689 ..LeuThrThrGlnTyrArgSerSerAla..... 697
1849 TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACT 1898


698 GlyValAsnPheAsnGly.....ValAsnGlyAsnMe 708
1899 GTTTTTC..... 1905
708 tSerPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProA 725
1906 AGCGGCAGACCGACACCGCACGCCTAC..... 1932
725 snGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsn 741
1933 AATCA 1937
742 IleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsnHi 758
1938 TTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCG 1987
758 sSerGlyArgGly.....AlaGluLeuLysMetSerGluIleA 771
1988 TGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTC... 2034
771 snIleSerAsn.....GlyAlaAsnPheThr 779
2035 CATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGC 2075
780 LeuAsnSerHisValArgGlyAspAspAlaPhe..... 790
2076 CAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTG 2125
791 .LysIleAsnLysAspLeuThrIleAsn..... 799
2126 GTGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACN 2175
800 AlaThrAsnSerAsnPhe... 805
2176 GGTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAAGTGATTGC 2225
805 805
2226 TTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACG 2275
806 .SerLeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaArgAsnA 822
2276 NTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCA 2325
822 laIleAsnSerThrTyrAsnIleSerIleLeuGlyGlyAsnValThrLeu 838
2326 AATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAA 2375
839 GlyGly.....GlnAsnSerSe 844
2376 CCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCC..... 2418
844 rSerSerIleThrGlyAsn.....IleThrIleGluLysAlaAlaAsnV 859
2419 ..ACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTAATCTA 2466
859 alThrLeuGluAlaAsnAsnAlaProAsnGlnGln.....AsnIle 872
2467 AGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAA 2516
873 ArgAspArgValIleLysLeuGlySerLeuLeuValAsnGlySerLeuSe 889
2517 GGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATA 2566
889 rLeuThrGlyGluAsnAlaAspIleLysGlyAsnLeuThrIleSerGluS 906
2567 AGGCAGTATTCCAT.....TTTGAAAACAGCCGCTTTACCGGACAA 2607
906 erAlaThrPheLysGlyLysThrArgAspThrLeuAsnIleThrGlyAsn 922
2608 CTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGAC 2657
923 PheThrAsnAsnGlyThrAlaGluIleAsnIleThrGlnGlyValValLy 939
2658 GCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCA 2707
939 sLeuGlyAsnValThrAsnAspGlyAspLeuAsnIle..... 951
2708 TTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGC 2757
951 951
2758 AGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGT 2807
952 ThrThrHisAlaLysArgAsnGlnArgSerIleIleGly.. 964
2808 TACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACG 2857
965 GlyAspIleIleAsnLysLysG 972
2858 GCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCG...GAACTCTTC 2904
972 lySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGly 988
2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTA 2954
989 GlyAsnIleSerGlnLys.....GluGlyAsnLe 998
2955 CACCTTGGCG.....GTCAACAATACCGGCAACGAACCCGTAAGCC 2995
998 uThrIleSerSerAspLysIleAsnIleThr..... 1008
2996 TCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCC... 3042
1009 ..LysGlnIleThrIleLysLysGlyIleAspGlyGluAspSerSerSer 1024
3043 GAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGC 3083
1025 AspAlaThrSerAsnAlaAsnLeuThrIleLys..... 1035
3084 CGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGC 3133
1035 1035
3134 ATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTC..... 3171
1036 ThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGly 1049
3172 ...GGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3218
1050 PheAsnLysAlaGluIle.....ThrAlaLysAsp..... 1059
3219 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAA 3268
1060 GlyArgAspLeu..... 1063
3269 GCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATG 3318
1064 ThrIleGlyAsnSerAsnAspGlyAsnSer 1073
3319 CAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTT 3368
1074 GlyAlaGluAlaLysThrValThrPheAsnAsnValLysAspSerLysIl 1090
3369 GGCGAAACAGCGCGAAGCGGAAACC.....CGGCCGGNTACCACCG 3409
1090 eSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerS 1107
3410 CCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCC 3459
1107 erSerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuThr 1123
3460 CAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGC...CGTTA 3506
1124 IleThrAlaLysAsnValGluValAsnLysAspIleThrSerLeuLysTh 1140

```
3507 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3556
1140 rValAsnIleThrAlaSerGluLysValThrThr.ThrAlaGlySerThr 1156

3557 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCN 3606
1157 IleAsnAlaThrAsnGlyLysAlaSerIleThrThrLysThrGlyAspIl 1173

3607 GTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGC...AAGA 3653
1173 eSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAspL 1190

3654 TTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGA 3703
1190 euThrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsn 1206

3704 AAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3753
1207 ValThrSerAlaThrGlyThrIleGlyGly.....ThrIleSerGl 1220

3754 GAAAACANCTTCGACGACGGCATCGGCA.....;.... 3781
1220 yAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnG 1237

3782 .....ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGC 3817
1237 lyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrGly 1253

3818 AAT......ACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCG 3861
1254 AsnThrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGl 1270

3862 GGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCG 3911
1270 nValAspLeuLeuAla......GlnAsnGlySerIleAlaGlySerIleA 1285

3912 CCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCG 3961
1285 snAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAla 1301

3962 GATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAA 4009
1302 GlySerAspIleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291

seq_documentation_block:

```
ID   AAW30291 standard; Protein; 1598 AA.
XX
AC   AAW30291;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW3.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW3; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 113
FT                   /note= "encoded by GTG"
FT   Misc-difference 864
FT                   /note= "encoded by TGT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90992.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 93-97; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW3 amino acid sequence was deduced from an
CC   isolated hmw3 gene (see AAT90992).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW4 (see AAW30292) have also been identified.  A
CC   conjugate comprising HMW3 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW3 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1598 AA;
```

alignment_scores:

Quality:	251.00	Length:	1460
Ratio:	0.369	Gaps:	69
Percent Similarity:	46.644	Percent Identity:	20.616

alignment_block:
US-09-303-518D-651 x AAW30291 ..

Align seg 1/1 to: AAW30291 from: 1 to: 1598

```
163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

213 TGAGGTNTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
174 eSerAsnGluAsnIleLysAlaArgAsnPheThrLeuGluGlnThrLysA 191

262 ..........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

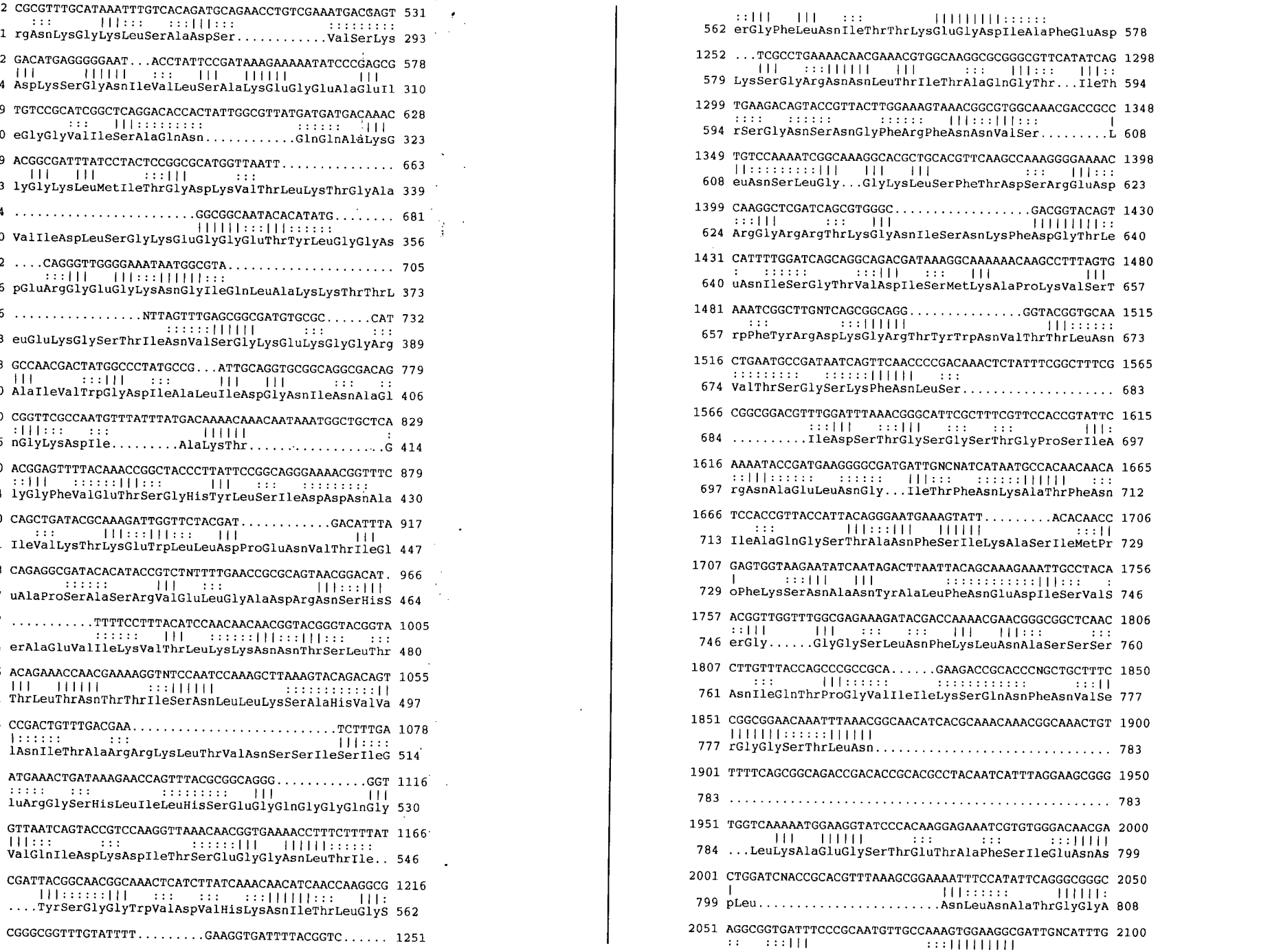
295 AACGGCGTGGCGGCATTGGTGGGC...........GATCAATATATTGT 332
208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT..... 375
224 eSerVal......AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

376 ........GCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAA 417
239 leThrIleSerAspIleIleAsnPro.........ThrIleThrTyrSer 252

418 ATTGTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGG 467
253 IleAla..............AlaProGluAsnGluAlaIleAsnLeuGl 264

468 CGAT.................................TANCATATGC 481
264 yAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleA 281
```



```
 808 snIleThrIle.....ArgGlnValGluGlyThrAspSerArg 820

2101 AGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATAC 2150
       |||                  |||||||||      ::::::   ||
 821 ValAsnLys.............GlyValAlaAlaLysLysAsnIleTh 832

2151 AATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAA 2200
     |:::       :::       |||
 832 rPheLysGlyGlyAsnIleThrPheGly..................... 841

2201 NCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACNGACNTNAGC 2250
                          ::::::   :::   |||:::   :::
 842 ......................SerGlnLysAlaThrThrGluIleLys 850

2251 GGCANTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGC 2300
     |||   |||::::::         :::   :::::::|||   |||
 851 GlyAsnValThrIleAsnLysAsnThrAsnAlaThrLeuArgGly..... 865

2301 NNCACTNAANGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCA 2350
               :::|||::::::   |||      :::       :::::::
 866 ..........AlaAsnPheAlaGluAsn...LysSerProLeuAsnIleA 878

2351 GCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGC......... 2391
     ::   |||:::   :::|||||||||||||:::   :::|||
 878 laGlyAsnValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIle 894

2392 ..................................AATGCCCAAGCA.. 2403
                                       |||   ||||||
 895 AsnIleAlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIl 911

2404 ..........ACATTTAATCAAGCCACATTAAACGGCAACNCATCGGNTT 2443
               |||||||||   |||            |||   ::::::|
 911 eThrAsnTyrThrPheAsnValAlaGlySerPheAspAsnAsnGlyAlaS 928

2444 CG..................GGCAATGCTTCATTTAATCTAAGCAACAAC 2475
     ||                 |||:::|||:::|||:::      ||||||
 928 erAsnIleSerIleAlaArgGlyGlyAlaLysPheLysAspIleAsnAsn 944

2476 GCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCT........... 2514
     :::::::::   :::      |||   ||||||::::::
 945 ThrSerSerLeuAsnIleThrThrAsnSerAspThrThrTyrArgThrIl 961

2515 ....AAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAG 2560
         |||:::|||:::|||::::::      :::|||:::::::::::::
 961 eIleLysGlyAsnIleSerAsnLys......SerGlyAspLeuAsnIleI 976

2561 CCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTC 2610
       ||||||                  :::|||           |||:::
 976 leAspLys.................LysSerAspAlaGluIleGlnIle 986

2611 AGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCT 2660
     :::|||:::         :::      |||:::::::::   |||::
 987 GlyGlyAsn............IleSerGlnLysGluGlyAsnLeuThrIl 999

2661 GCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTA 2710
     :   |||              :::::::|||:::   |||      ||||
 999 eSerSer............AspLysValAsnIleThrAsn...GlnIleT 1011

2711 CACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGN 2760
     ||::::::                    ||||||::::::   |||
1011 hrIleLys...................AlaGlyValGluGlyGly... 1019

2761 GTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTAC 2810
                                       ||||||      |||   ::
1020 ............................ArgSerAspSerSerGluAl 1026

2811 ACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCA 2860
     :     :::::::::      :::   :::   |||   :::   |||
1026 aGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLeuAlaGlyA 1043

2861 AATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2910
       ||||||   :::|||      :::      :::|||:::   :::
1043 spLeuAsnIleSerGlyPheAsnLys......AlaGluIleThrAlaLys 1057

2911 CGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTT 2960
     :::::::::   |||   :::::::::::|||   |||
1058 AsnGlySerAspLeuThrIleGlyAsnAlaSerGlyGly..... 1070

2961 GGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGG 3010
               |||:::      :::   |||::::::|||::::::
1071 .........AsnAlaAspAlaLysLysValThrPheAspLysVal.... 1082

3011 TAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3060
                  ||||||:::|||      :::::::   |||   |||
1083 ...........LysAspSerLysIleSerThrAspGlyHisAsnValThr 1095

3061 CTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCG 3110
     |||:::::::|||
1096 LeuAsnSerGlu..................................... 1099

3111 CAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTT 3160
                                       ||||||   ::::::   |
1100 ..............................ValLysThrSerAsnGlyS 1106

3161 CCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3210
     ||:::::   |||                              :::||||||
1106 erSerAsnAlaGly..........................AsnAspAsn 1113

3211 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAA 3260
     :::   :::|||
1114 SerThrGlyLeu..................................... 1117

3261 GACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCG 3310
                                       :::   ::::::::::::|||
1118 ............................ThrIleSerAlaLysAspValT 1125

3311 GCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3360
       :::   :::          :::   |||   ::::::   :::
1125 hrValAsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaAla 1141

3361 AGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGC 3410
     ::::::   :::      :::|||:::   |||      :::||||||::
1142 AlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThrGl 1158

3411 CTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCC 3460
     : |||         :::|||      |||      :::::::::
1158 y.SerValGluValThrAlaGlnAsnGlyThrIleLysGlyAsnIleThr 1174

3461 AACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCC 3510
     ::::::|||::::::         :::||||||         |||
1175 SerGlnAsnValThrVal......ThrAlaThrGluAsnLeuValThrTh 1189

3511 AATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGT 3560
            :::   :::|||                :::::   :::|||
1189 rGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer.... 1204

3561 ACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTT 3610
                                       ||||||:::
1205 ........................ThrLysThrGlyAspIleLysGly 1212

3611 GGA...CAAGCNGCATCCGGNACACCAAACACTACCGTTCGC........ 3649
     |||   :::::::::||||||:::   |||   |||:::
1213 GlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnTh 1229

3650 ....AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGG 3695
       |||:::|||      |||::::::      |||   :::   ::::
1229 rLeuLysValSerAsnIleThrGlyGlnAspValThrValThrAlaAspA 1246

3696 TATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACA 3745
     ::         |||:::::::||||||::::::   |||:::      |||
1246 laGlyAlaLeuThrThrThrAlaGlySerThrIleSerAla......Thr 1260

3746 ACCGGA.............................................. 3751
     ||||||
1261 ThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysVa 1277
```

```
     3752 ....CCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGC 3797
                  ::::::   |||:::|||      ||||||         |||
     1277 lGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuA 1294

     3798 CCACGGCGCCGTTTTCGGGCAAT.............ACGGCATCGGCAGGT 3835
            ::::::   :::||||||||||            ||||||:::::|||
     1294 laValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGly 1310

     3836 TCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCA 3885
          :::   ::::::::::::   :::   :::   ::::::   :::
     1311 LysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSerValTh 1327

     3886 GACGGCATCGGAGGCAAAA.................TCCGCCGCCGCGT 3917
           |||:::|||::::::                    |||:::   :::
     1327 rThrSerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsnThrV 1344

     3918 GCTGCATTACGGCATTCAGGCACGAT...ACCGCGCCGGTTTCGGCGGAT 3964
               :::||||||          |||   |||         ||||||
     1344 alAsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLys 1360

     3965 TCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4014
               :::::::::   :::|||      :::   :::|||   |||
     1361 ValGluAlaLysAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLe 1377

     4015 GATTACCGCTACGAAAACGTCAATATCG 4042
               |||   |||   :::|||||||:::
     1377 uThrThrGlnThrGlySerSerIleThr 1386
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294

seq_documentation_block:

```
ID   AAW30294 standard; Protein; 1477 AA.
XX
AC   AAW30294;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW2.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW2; hmw2A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 34
FT                   /note= "encoded by TTC"
FT   Misc-difference 35
FT                   /note= "encoded by CGC"
FT   Misc-difference 36
FT                   /note= "encoded by TAT"
FT   Misc-difference 37
FT                   /note= "encoded by GTT"
FT   Misc-difference 38
FT                   /note= "encoded by ACT"
FT   Misc-difference 39
FT                   /note= "encoded by ATC"
FT   Misc-difference 40
FT                   /note= "encoded by TTT"
FT   Misc-difference 41
FT                   /note= "encoded by AGG"
FT   Misc-difference 42
FT                   /note= "encoded by TGT"
FT   Misc-difference 43
FT                   /note= "encoded by AAC"
FT   Misc-difference 424
FT                   /note= "encoded by TAT"
FT   Misc-difference 426
FT                   /note= "encoded by TCC"
FT   Misc-difference 428
FT                   /note= "encoded by GAC"
FT   Misc-difference 429
FT                   /note= "encoded by AGC"
FT   Misc-difference 434
FT                   /note= "encoded by AAA"
FT   Misc-difference 435
FT                   /note= "encoded by ACA"
FT   Misc-difference 442
FT                   /note= "encoded by CCT"
FT   Misc-difference 444
FT                   /note= "encoded by GAT"
FT   Misc-difference 446
FT                   /note= "encoded by ACA"
FT   Misc-difference 448
FT                   /note= "encoded by GAA"
FT   Misc-difference 454
FT                   /note= "encoded by CGC"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90995.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 73-78; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW2 (123 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW2 amino acid sequence was deduced from the
CC   hmw2 gene sequence (see AAT90995 and AAT90997).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW1 (see AAW30293), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW2
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW2 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1477 AA;
```

alignment_scores:

```
            Quality:  250.50              Length:   1399
              Ratio:  0.403                 Gaps:     66
 Percent Similarity:  44.389   Percent Identity: 19.085
```

alignment_block:

US-09-303-518D-651 x AAW30294 ..

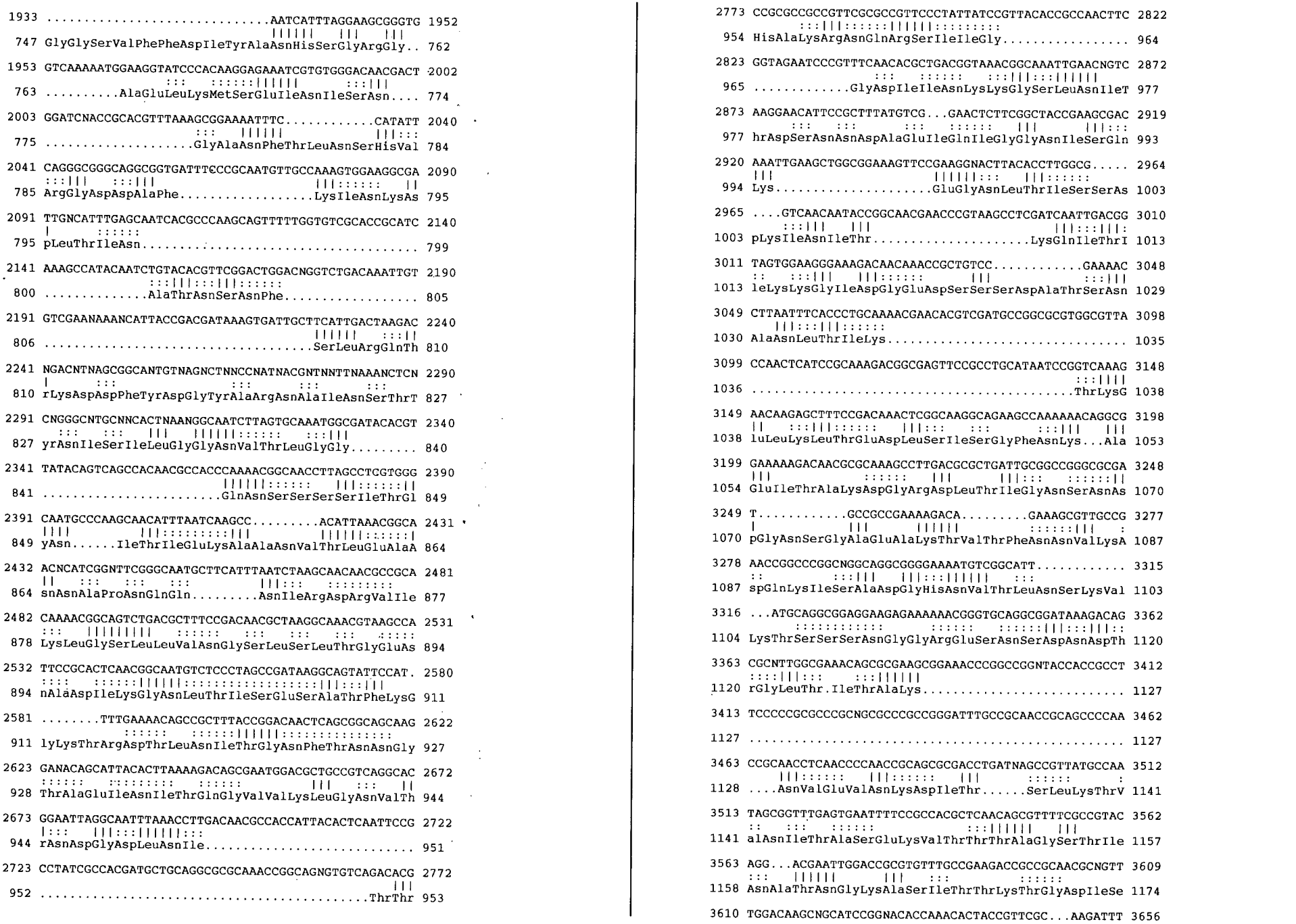
Align seg 1/1 to: AAW30294 from: 1 to: 1477

```
     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACAA 227
         |||:::||||||:::::::::          |||             :::||
     282 AsnGlnGlyLysLeuSerAla.........AspSerValSerLysAspLy 295

     228 AAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATT 277
         |:::|||::::::|||   ::::::              |||
     295 sSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleGlyG 312

     278 TTTCTGTGGTGTCGCGTAAC..............GGCGTGGCGGCATTG 312
```





```
      |||      ::::::::::|||   ::::::::::::          ::::
 1174 rGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAspLeuT 1191

 3657 CCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAA 3706
      :::::   :::::::::|||      :::   ::::::   ::::::
 1191 hrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsnVal 1207

 3707 ACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3756
      ||||||||||:::|||:::   ::::::          |||   :::   ::
 1208 ThrSerAlaThrGlyThrIleGlyGly.........ThrIleSerGlyAs 1221

 3757 AACANCTTCGACGACGGCATCGGCA..................... 3781
      :|||   ::::::||||||:::|||
 1221 nThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyA 1238

 3782 ..........ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT 3820
                 :::   :::|||      |||         :::||||||
 1238 laGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsn 1254

 3821 ......ACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT 3864
            |||::::::||||||||||:::   :::|||:::::::::::||
 1255 ThrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnVa 1271

 3865 TTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCG 3914
      |        |||      |||::::::|||   |||   |||      |
 1271 lAspLeuLeuAla......GlnAsnGlySerIleAlaGlySerIleAsnA 1286

 3915 CGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGAT 3964
      ||:::   :::|||               |||      :::   |||
 1286 laAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGly 1302

 3965 TCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAA 4009
      |||      :::   ||||||:::      :::|||:::::::|||
 1303 SerAspIleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG30355

seq_documentation_block:

```
ID   ABG30355 standard; Protein; 1606 AA.
XX
AC   ABG30355;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #30346.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS94542.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 60714; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1606 AA;
```

alignment_scores:

```
            Quality:  242.50              Length:    739
              Ratio:  0.691                 Gaps:     31
 Percent Similarity:  47.497   Percent Identity: 19.892
```

alignment_block:

US-09-303-518D-651 x ABG30355 ..

Align seg 1/1 to: ABG30355 from: 1 to: 1606

```
 2311 GGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGC 2360
      |||:::|||::::::|||   :::   :::   |||
  632 GlyThrLeuThrValAsnAspAsnGlyLysAlaThr..............  643

 2361 CACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2410
                          ::::::|||   |||:::   |||:::::::
  644 ..................AspIleValGlnAsnSerGlyAlaAlaLeuG  654

 2411 ATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAAT......... 2451
      ::   :::|||   ||||||      :::   ||||||:::
  654 lnThrSerThrAlaAsnGly...IleGluIleSerGlyThrHisGlnTyr  669

 2452 GCTTCATTTAATCTAAGCAACAACGCCGCACAAAAC.............. 2487
      ::::::|||::::::|||:::|||   |||   |||
  670 GlyThrPheSerIleSerGlyAsnLeuAlaThrAsnMetLeuLeuGluAs  686

 2488 ....GGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATT 2533
          |||:::|||   :::      ::::::::::|||:::   |||
  686 nGlyGlyAsnLeuLeuValLeuAlaGlyThrGluAlaArgAspSerThrV  703

 2534 CCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTT 2583
        :::   :::|||                        |||:::   ::::::
  703 alGlyLysGlyGly.....................AlaMetGlnAsnLeu  712

 2584 GAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATT 2633
        ::::::      |||         ||||||::::::   |||
  713 GlyGlnAspSerAlaThrLysValAsnSerGlyGlyGlnTyrThrLeuGl  729

 2634 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA.... 2679
       :::   ||||||:::      :::|||      :::   :::|||
  729 yArgSerLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnV  746

 2679 ..................................................  2679

  746 alAlaGlyGlyThrAlaIleValTyrAlaGlyThrLeuAlaAspAlaSer  762

 2680 ..............GGCAATTTAAACCTT..................... 2694
```

```
                    |||:::|||:::|||
 763 ValSerGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnVa 779

2695 ....................................GACAACGCCACCA 2707
                                         |||:::|||||||:
 779 lThrProValLysLeuGluGlyAlaValArgIleThrAspSerAlaThrL 796

2708 TTACACTCAATTCCGCCTAT.........CGCCACGATGCTGCAGGCGCG 2748
     ::|||||||:::::::::                |||   :::::::|||
 796 euThrLeuGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAla 812

2749 CAAACCGGCAGNGTG.......................TCAGACACGCC 2774
     :::   |||:::|||                      :::   |||
 813 SerArgGlySerValTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSe 829

2775 GCGCCGCCGTTCGCGC...CGTTCCCTATTA..................... 2802
      :::   :::   |||   :::||||||||||
 829 rAsnCysGluTyrArgValAsnSerLeuLeuLeuAsnAspGlyAspValT 846

2803 ...........TCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAAC 2841
             :::::::::   |||:::::::::::::::   :::|||
 846 yrLeuSerAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTyrAsn 862

2842 ACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTAT 2891
     |||||||||:::|||   :::|||:::   :::|||:::|||   :::
 863 ThrLeuThrThrAsn...GluLeuSerGlySerGlyAsnPheTyrLeuHi 878

2892 GTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTT 2941
      :::::::::   |||   |||:::|||:::|||   :::   :::::::
 878 sThrAsnValAlaGlySerArgGlyAspGlnLeuValValAsnAsnAsnA 895

2942 CCGAAGGNACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTA 2991
     ::   |||::::::   :::   |||:::::||||||   :::|||
 895 laThrGlyAsnPheLysIlePheValGlnAspThrGlyValSerProGln 911

2992 AGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTC 3041
     |||   |||   :::|||:::   ||||||
 912 SerAspAspAlaMetThrLeuGlyGluGly.................Me 922

3042 CGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGT 3091
         :::|||:::::::::::::       ||||||   |||::::
 922 tLeuArgLeuArgTrpAlaIleProAlaValSerValAspLeuGlyThrT 939

3092 GGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCG 3141
     :::::|||   :::::::::::::||||||  :::|||
 939 yrGluTyr...ValLeuLysSerAspGly.ThrSer.............. 949

3142 GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAA 3191

 949 ..................................................... 949

3192 ACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCG 3241
                     :::   :::|||||||::   ::::::|||
 950 ....................AsnTrpThrLeuThrAsnAspValLysPro 959

3242 GGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNG 3291
     :::     |||                  :::||||||
 960 AsnProAspPro.................IleProAsn............ 966

3292 GCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACG 3341

 966 ..................................................... 966

3342 GGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAA 3391
                                                      |||
 967 ..........................................ProLys 968

3392 CCCGGCCGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGAT 3441
     |||         |||   |||:::|||   |||         :::
 969 ProAspProLysProAspProLysProAspPro..AsnProLys...... 982

3442 TTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGA 3491
        |||:::|||   |||:::|||   |||   |||   |||::::::
```

```
 983 ...ProAspProThrProAspProThrProThrProValProGluLysAr 998

3492 CCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGC 3541
       ::::::         :::::::::::::::   :::::::::|||||||
 998 gIleThrPro......SerThrAlaAlaValLeuAsnMetAlaAlaThrL 1013

3542 TCAACAGCGTTTTC..............GCCGTACAGGACGAATTGGAC 3576
     ||      ||||||               :::::::::::::::|||:::
1013 euProLeuValPheAspAlaGluLeuAsnSerIleArgGluArgLeuAsn 1029

3577 CGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGCATCCG 3626
        :::   |||:::   ::::::|||   ||||||   ::::::   ::
1030 IleMetLysAlaSerProAsnAsnAsnAsnValTrpGlyAlaThrTyrAs 1046

3627 GNACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAA 3676
     ::::   ::::::       ::::::          :::|||:::|||
1046 nThrArgAsnAsnValThrThrAspAlaGlyAlaGlyTyrGluGlnThrL 1063

3677 CCGACCTGCGCCAAATCGGTATGCAGAAAAAC............CTCGGC 3714
                  :::|||::::::::::::              |||
1063 euThrGlyMetThrValGlyIleAspSerArgAsnAspIleProGluGly 1079

3715 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTT 3764
               :::|||   :::::::::::::::::::::::::   :::||
1080 IleThrThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIleSerPh 1096

3765 CGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCG 3814
     ||||   |||                        :::||||||:::|||   |
1096 eAspArgGly.....................GlyHisGlySerVal...G 1105

3815 GGCAATACGGCATCGGC................................ 3831
     ||:::|||::::::|||
1105 lySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerGlyPhe 1121

3832 ........................AGGTTCGACATCGGCATC..... 3849
                             ||||||      ::::::
1122 TyrLeuAspGlyValValLysLeuAsnArgPheLysSerAsnValAlaGl 1138

3850 .......AGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCA 3892
           |||:::|||::::::   :::::::::      |||:::|||:
1138 yLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyL 1155

3893 TCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGA 3942
     ::||||||                                |||::::::
1155 euGlyGlyHis.........................IleGluThrGly 1162

3943 TACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAAC 3992
        |||         |||:::::::::::::   ||||||   :::   ||
1163 MetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuTh 1179

3993 GCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAAC.......... 4032
     |   :::   :::::::::   :::|||:::   :::|||
1179 rGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMetLysS 1196

4033 ..GTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGN...GCG 4077
          ::::::   |||   ::::::            ||||||
1196 erLysSerValAspThrArgSerIle............TyrArgGluLeu 1208

4078 GGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATNTCCATCAC 4127
     |||          :::|||::::::::::   ::::::      ::::::
1209 GlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluValGl 1225

4128 NCCTTATTTNAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAA 4177
      |||:::   :::   :::                              |||:
1225 uProTrpLeuLysAlaAlaValArgLysGluPheValAspAspAsnArgV 1242

4178 CACGCGTCAATACC 4191
     :::::||||||:::
1242 alLysValAsnSer 1246
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143

seq_documentation_block:

```
ID   AAU34143 standard; Protein; 2086 AA.
XX
AC   AAU34143;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #419.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52002.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5639; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2086 AA;


alignment_scores:
            Quality:  241.50              Length:   1531
              Ratio:  0.325                Gaps:     71
 Percent Similarity:  48.465   Percent Identity: 20.248

alignment_block:
 US-09-303-518D-651 x AAU34143   ..

 Align seg 1/1  to: AAU34143  from: 1  to: 2086
```

```
231 AGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTT 280
    :::|||   |||:::   |||:::|||::::::   :::||| ||||||:
 28 GlnGlySerGlyGluThrIleSerAspGluGlnArgSerAsp.AspPheT 44

281 CTGTGGTGTCG..................................... 291
    ::|||||:::
 44 hrValValAlaProGlnProAsnHisAlaAsnThrLysIleTrpGlnAsn 60

292 .....CGTAACGGCGTGGCGGCATTGGTGGGCGA 320
                  :::|||      :::   |||:::::
 61 GlyHisIleAspIleThrProAsnAsnProSerGlyHisLeuIleAsnPr 77

321 TCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATT 370
              :::::::::|||:::::        :::|||
 77 oThrGlnAlaMetAspIleAlaTyrThrGluLysMetGlyAsn..... 91

371 TTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420
      |||||||||                |||      :::   ::::::
 92 ..GlyAlaGlu.................HisSerLysThrIleAsnVal 101

421 GTGAAAAGAAATAAT.................TATAAGCCTGACAATTC 452
    |||:::   :::|||                 |||||||||
102 ValArgGlyGlnAsnAsnGlnTrpThrIleAlaAsnLysProAsp..... 116

453 ACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCA 502
                                                ::::||||
117 .....TyrValT 119

503 CA...GATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATACC 549
    ||   ||||||:::      :::::::|||      :::::::::||||||
119 hrLeuAspAlaHisThrGlyLysValThr......PheAsnAlaAsnThr 133

550 TATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACA 599
      :::   ::::::                 :::   |||:::|||||
134 ValLysProAsnSerAlaIleThrIleThrProLysAlaGlyThrGlyHi 150

600 CCACTATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCG 649
    |                                  :::   |||   :::
150 s.............................SerAlaSerSerAsnP 156

650 GCGCATGGTTAATTGGCGGCAATACACAT..................... 678
      :::   |||   :::      ||||||
156 roSerThrLeuThrAlaProAlaThrHisThrValAsnThrThrGluIle 172

679 ATGCAGGGTTGGGGAAATAAT...GGCGTANTTAGTTTGAGCGGCGATGT 725
    ::::::   :::|||:::|||      ::::::::::::::::::   ||
173 ValLysAspTyrGlySerAsnValThrAlaAlaGluIleAsnAsnAlaVa 189

726 GCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCG 775
    |:::   ||||||                                |||:::
189 lGlnValAlaAsnLys.................ArgThrAlaThrIleL 200

776 ACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTG 825
      :::|||:::   |||                ||||||:::   |||
200 ysAsnGlyThrAlaMet............ProThrAsn.LeuAlaGlyGl 212

826 CTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGG 875
     ||||||                 |||   ||||||:::      |||
212 ySerThr.................ThrThrIleProValThrValThrT 223

876 TTTCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCG 925
      ::::::               |||||||||            |||:::
223 yrAsnAsp..............GlySerThrGluGluValGlnGluSer 234

926 ATACACATACCGTCTNTTTTGAACCGCGCAGTAACGGA.......CATTT 968
    |||         :::      ::::::   :::|||         |||::
235 IlePheThrLysAlaAspLysArgGluLeuIleThrAlaLysAsnHisLe 251

969 TTCCTTTACATCCAACAACAACGGT......ACGGGTACGGTAACAGAAA 1012
    ::::         :::::::::|||         ||||||:::|||:::
251 uAspAspProValSerThrAspGlyLysLysProGlyThrIleThrGlnT 268

1013 CCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTG 1062
```

```
      |||:::   :::   |||   :::   :::::::::|||::::::
  268 yrAsnAsnAlaIleHisAsnAlaGlnGlnGlnIleAsnThrAlaLysThr 284

 1063 ...TTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACGCGGC 1109
        ::::::   :::|||:::          |||       :::::
  285 GluAlaGlnGlnValIleAsnAsnGluArgAlaThrProGlnGlnValSe 301

 1110 AGGGGGTGTTAATCAGTACCGT.........CCAAGGTTAAACAACGGTG 1150
      :   ::::::::::::   |||       ::::::|||:::::::
  301 rAspAlaLeuThrLysValArgAlaAlaGlnThrLysIleAsnGluAlaL 318

 1151 AAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAAC 1200
      ::   |||                |||::::::|||:::   |||:::
  318 ysAlaLeuLeuGlnAsnLysGluAspAsnSerGlnLeuValThrSerLys 334

 1201 AAC..............ATCAACCAA..................... 1212
      |||              :::|||||||
  335 AsnAsnLeuGlnSerSerValAsnGlnValProSerThrThrGlyMetTh 351

 1213 ....................GGCGCGGGCGGTTTGTATTTTGAAGGTG 1240
                          :::|||          |||   :
  351 rGlnGlnSerIleAspAsnTyrAsnAlaLysLysArgGluAlaGluThrG 368

 1241 ATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTT 1290
      :::::|||::::::   :::      ::::::   ::::::::::::
  368 luIleThrAlaAlaGlnArgValIleAspAsnGlyAspAlaThrAlaGln 384

 1291 CATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAA 1340
      :::|||||:::::::::|||::::::   :::              ||
  385 GlnIleSerAspGluAsnThrThrAlaGlnArg.............As 396

 1341 CGACCGCCTGTCCAAAATCGGCAAAGGCACG......CTGCACGTTCAA. 1383
      ||||      :::::::|||:::::::::|||           |||:::
  396 nAspLeuThrAsnGlnIleSerGlnAlaThrAsnLeuAlaAlaValGluS 413

 1384 ............................GCCAAAGGGGAAAACCAA 1401
                                 |||   |||:::   |||
  413 erValLysGlnSerAlaAsnSerLeuAspGlyAlaMetGlyAsnLeuGln 429

 1402 GGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGA 1451
         :::|||:::      :::|||||||:::      ::::::      ||
  430 ThrAlaIleAsnAspLysSerGlyThrLeuAlaSerGlnAsnPheLeuAs 446

 1452 C...GATAAAGGCAAAAAACAAGCCTTTAGTGAA.............. 1482
      |   |||:::   |||::::::|||:::::::::
  446 pAlaAspGluGlnLysArgAsnAlaTyrAsnGlnAlaValSerAsnAlaG 463

 1483 ........ATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCC 1524
              ::::::      :::|||   :::|||::::::   :::
  463 luThrIleLeuAsnLysGlnThrGlyProAsnThrAlaLysThrAlaVal 479

 1525 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1574
      ::::::   :::|||
  480 GluGlnAlaLeuAsn................................ 484

 1575 TTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCG 1624
           ::::::|||:::   :::   :::::::::      ||||||
  485 ....AsnValAsnSerAlaLysHisAlaLeuAsnGlyThrGlnAsn.... 498

 1625 ATGAAGGGGCGATGATTGNCNATCATAATGCCACAACAACATCCACCGTT 1674
                           :::|||||||      ::::::   :::
  499 ...................LeuAsnAsnAlaLysGlnAlaAlaIleThr 508

 1675 ACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATCAA 1724
      :::|||:::|||   :::::::::::::|||   :::         :::::
  509 AlaIleAsnGlyAlaSerAspLeuAsnGlnLysGlnLysAspAlaLeuLy 525

 1725 TAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1774
      :           :::   :::   :::   |||:::         ::::
  525 sAlaGlnAlaAsnGlyAlaGlnArgValSerAsnAlaGlnAspValGlnA 542

 1775 AAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCC 1824
      ::::::::|||:::   |||      ::::::   :::   :::   |||
```

```
  542 rgAsnAlaThrGluLeuAsnThrAlaMetGlyThrLeu...LysHisAla 557

 1825 GCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAA 1874
           |||:::|||      |||:::::::::      :::|||:::::
  558 IleAlaAspLysThrAsnThrLeuAlaSerSerLysTyrValAsnAlaAs 574

 1875 CATCACGCAA.......................ACAAACGGCAAACTGT 1900
      :   |||:::                       ||||||::::::   :
  574 pSerThrLysGlnAsnAlaTyrThrThrLysValThrAsnAlaGluHisI 591

 1901 TTTTCAGCGGCAGACCGACA.......................CCGCAC 1926
      :::::||||||   ||||||
  591 leIleSerGlyThrProThrValValThrThrProSerGluValThrAla 607

 1927 GCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACA 1976
      |||   |||::::::::::|||:::   ::::::::::::|||      ::
  608 AlaAlaAsnGlnValAsnSerAlaLysGlnGluLeuAsnGlyAspGluAr 624

 1977 AGGAGAAATCGTGTGGGACAAC..............GACTGGATCNACC 2011
      :   :::::::::   :::|||                |||   :::::::
  624 gLeuArgValAlaLysGlnAsnAlaAsnThrAlaIleAspAlaLeuThrG 641

 2012 GC......ACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCG 2055
      ::      |||   :::      ::::::      |||   |||||||||
  641 lnLeuAsnThrProGlnLysAlaLysLeuLysGluGlnValGlyGlnAla 657

 2056 GTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGNCATTTGAGCAA 2105
               :::::::::      |||:::   :::   :::      ||
  658 AsnArgLeuGluAspIleGlnThrValGlnThrAsnGlyGlnAlaLeuAs 674

 2106 TCACGCC.......................................... 2112
      |:::|||
  674 nAsnAlaMetLysGlyLeuArgAspSerIleAlaAsnGluThrThrValL 691

 2113 ........CAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATC 2154
              |||            :::::::|||:::::::::   |||
  691 ysAlaSerGlnAsnTyrThrAspAlaSerProAsnAsnGlnSerThrTyr 707

 2155 TGTACACGT...TCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAAN 2201
         :::      |||:::      |||:::   |||   ::::::
  708 AsnSerAlaValSerAsnAlaLysGlyIleIleAsnGlnThrAsnAsnPr 724

 2202 CATTACCGACGATAAAGTGATTGCTTCATTGACT.............. 2235
              |||   :::::::|||::::::   |||
  724 oThrMetAspThrSerAlaIleThrGlnAlaThrThrGlnValAsnAsnA 741

 2236 ..AAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNTTNA 2283
        |||:::      :::|||:::      |||      :::      :::
  741 laLysAsnGlyLeuAsnGlyAlaGluAsnLeuArgAsnAlaGlnAsnThr 757

 2284 AAN.........CTCNCNGGGCNTGCNNCACTNAANGGCAAT........ 2316
      :::          |||          :::   |||:::::::|||
  758 AlaLysGlnAsnLeuAsnThrLeuSerHisLeuThrAsnAsnGlnLysSe 774

 2317 ....CTTAGTGCAAATGGCGAT..............ACACGTTATACAG 2347
          :::|||::::::   |||               ::::::   |||:
  774 rAlaIleSerSerGlnIleAspArgAlaGlyHisValSerGluValThrA 791

 2348 TCAGCCACAACGCC..............ACCCAAAACGGCAACCTTAGC 2382
      :::::   ||||||              ||||||   ||||||||||:::
  791 laAlaLysAsnAlaAlaThrGluLeuAsnThrGlnMetGlyAsnLeuGlu 807

 2383 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2432
         :::   :::   |||   |||   :::|||::::::         ||
  808 GlnAlaIleHisAspGlnAsnThrValLysGlnGlyVal.........As 821

 2433 CNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCAC 2482
      |   :::   :::   :::|||:::   :::      :::||||||::::
  821 nPheThrAspAlaAspLysAlaLysArgAspAlaTyrThrAsnAlaValS 838

 2483 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGC... 2529
      ::::::::::          |||:::   :::      ||||||:::|||
  838 erArgAlaGluThrIleLeuAsnLysThrGlnGlyAlaAsnThrSerLys 854
```



```
        :|||   ::::::   :::   ::::::   |||
    1394 yLeuProAsnLeuThrGlnLeuGlnLysAspAsn 1405

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017

seq_documentation_block:
ID   AAU37017 standard; Protein; 5795 AA.
XX
AC   AAU37017;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1187.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54876.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12610; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5795 AA;


alignment_scores:
           Quality:  241.50            Length:   1531
             Ratio:  0.325               Gaps:     71
 Percent Similarity:  48.465   Percent Identity: 20.248

alignment_block:
 US-09-303-518D-651 x AAU37017   ..

 Align seg 1/1  to: AAU37017  from: 1  to: 5795

     231 AGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTT 280
         :::|||   |||:::   |||:::|||::::::   :::||| ||||||:
    2305 GlnGlySerGlyGluThrIleSerAspGluGlnArgSerAsp.AspPheT 2321

     281 CTGTGGTGTCG..................................... 291
         ::||||||:::
    2321 hrValValAlaProGlnProAsnHisAlaAsnThrLysIleTrpGlnAsn 2337

     292 .....................CGTAACGGCGTGGCGGCATTGGTGGGCGA 320
                              :::|||      :::   |||::::::
    2338 GlyHisIleAspIleThrProAsnAsnProSerGlyHisLeuIleAsnPr 2354

     321 TCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATT 370
                 ::::::::::|||::::::         :::|||
    2354 oThrGlnAlaMetAspIleAlaTyrThrGluLysMetGlyAsn....... 2368

     371 TTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420
           |||||||||                  |||      :::   ::::::
    2369 ..GlyAlaGlu.................HisSerLysThrIleAsnVal 2378

     421 GTGAAAAGAAATAAT.................TATAAGCCTGACAATTC 452
         |||:::   :::|||                 |||||||||
    2379 ValArgGlyGlnAsnAsnGlnTrpThrIleAlaAsnLysProAsp..... 2393

     453 ACACCCTTACAACGGCGATTANCATATGCCGCGTTTGCATAAATTTGTCA 502
                                                    :::||||
    2394 .....................................TyrValT 2396

     503 CA...GATGCAGAACCTGTCGAAATGACGAGTGACATGAGGGGGAATACC 549
         ||   ||||||:::      :::::|||      :::::::::||||||
    2396 hrLeuAspAlaHisThrGlyLysValThr......PheAsnAlaAsnThr 2410

     550 TATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACA 599
         :::   :::::                 :::   |||:::|||||
    2411 ValLysProAsnSerAlaIleThrIleThrProLysAlaGlyThrGlyHi 2427

     600 CCACTATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCG 649
         |                              :::   |||   :::
    2427 s.............................SerAlaSerSerAsnP 2433

     650 GCGCATGGTTAATTGGCGGCAATACACAT..................... 678
           :::   |||   :::      ||||||
    2433 roSerThrLeuThrAlaProAlaThrHisThrValAsnThrThrGluIle 2449

     679 ATGCAGGGTTGGGGAAATAAT...GGCGTANTTAGTTTGAGCGGCGATGT 725
         ::::::   :::|||:::|||      :::::::::::::::::   ||
    2450 ValLysAspTyrGlySerAsnValThrAlaAlaGluIleAsnAsnAlaVa 2466

     726 GCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCG 775
         |:::   ||||||                              |||:::
    2466 lGlnValAlaAsnLys.................ArgThrAlaThrIleL 2477

     776 ACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTG 825
           :::|||:::   |||              ||||||  :::   |||
    2477 ysAsnGlyThrAlaMet............ProThrAsn.LeuAlaGlyGl 2489

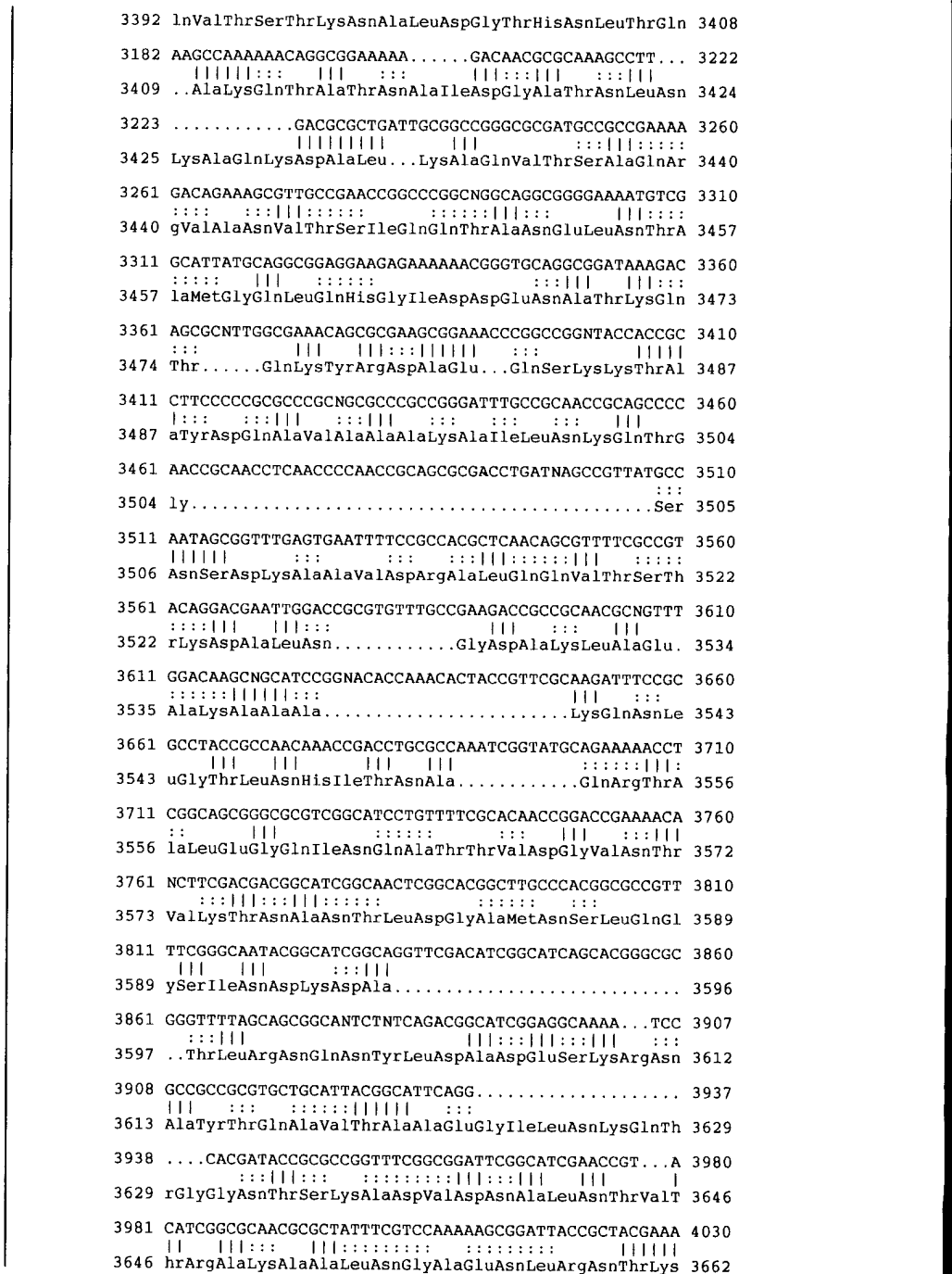
     826 CTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGG 875
          ||||||                  |||   |||||||::      |||
    2489 ySerThr.................ThrThrIleProValThrValThrT 2500

     876 TTTCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCG 925
           ::::::              |||||||||            |||:::
    2500 yrAsnAsp..............GlySerThrGluGluValGlnGluSer 2511

     926 ATACACATACCGTCTNTTTTGAACCGCGCAGTAACGGA.......CATTT 968
         |||       :::     ::::::   :::|||          |||::
    2512 IlePheThrLysAlaAspLysArgGluLeuIleThrAlaLysAsnHisLe 2528
```




3098 nPheThrAspAlaAspLysAlaLysArgAspAlaTyrThrAsnAlaValS 3115
2483 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGC... 2529
3115 erArgAlaGluThrIleLeuAsnLysThrGlnGlyAlaAsnThrSerLys 3131
2530 CATTCCGCACT 2540
3132 GlnAspValGluAlaAlaIleGlnAsnValThrSerAlaLysAsnAlaLe 3148
2541 CAACGGC......AATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTG 2584
3148 uAsnGlyAspGlnAsnValThrAsnAlaLysAsnThrAlaLysHisAlaL 3165
2585 AAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATTA 2634
3165 euAsnAsn......LeuThrSerIleAsnAsnAlaGlnLysArgAspLeu 3179
2635 CACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAA..... 2676
3180 ThrThrLysIleAspGlnAlaThrThrValAlaGlyValGluAlaValSe 3196
2677 TTAGGCAATTTA......AACC 2692
3196 rAsnThrGlyThrGlnLeuAsnThrAlaMetAlaAsnLeuGlnAsnGlyI 3213
2693 TTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCA 2742
3213 leAsnAspLysAlaAsnThrLeuAlaSerGluAsnTyrHisAspAlaAsp 3229
2743 GGC.....GCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCG 2780
3230 SerAspLysLysThrAlaTyrThrGlnAlaValThrAsnAlaGluAsnIl 3246
2781 CCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAAT 2830
3246 eLeuAsnLysAsnSer...GlySerAsnLeuAspLysAlaAlaValGluA 3262
2831 CCCGTTTCAACACGCTGACG.....GTAAACGGC..... 2859
3262 snAlaLeuSerGlnValThrAsnAlaLysGlyAlaLeuAsnGlyAsnHis 3278
2860 AAATTGAACNGTCAAGGAACATTCCGCTTTATGTC 2894
3279 AsnLeuGluGlnAlaLysSerAsnAlaAsnThrThrIleAsnGlyLeuGl 3295
2895 GGAACTCTTCGGCTACCGAAGCGACAAATTGAAG..... 2928
3295 nHisLeuThrThrAlaGlnLysAspLysLeuLysGlnGlnValGlnGlnA 3312
2929 CTGGCGGAAAGTTCCGAAGGNACTTAC 2955
3312 laGlnAsnValAlaGlyValAspThrValLysSerSerAlaAsnThrLeu 3328
2956 ACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATT 3005
3329 AsnGlyAlaMetGlyThrLeuArgAsn.....SerIleGlnAspAs 3342
3006 GACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATT 3055
3342 nThrAlaThrAsnAsnGlyGlnAsnTyrLeuAspAlaThrGluSerAsnL 3359
3056 TCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTC 3105
3359 ysThrAsnTyrAsnAsnAlaValAspSer.AlaAsnGlyValIleAsnAl 3375
3106 ATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGG..... 3142
3375 aThrSerAsnProAsnMetAspAlaAsnAlaIleAsnGlnIleAlaThrG 3392
3143 TCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3181


3392 lnValThrSerThrLysAsnAlaLeuAspGlyThrHisAsnLeuThrGln 3408
3182 AAGCCAAAAAACAGGCGGAAAAA......GACAACGCGCAAAGCCTT... 3222
3409 ..AlaLysGlnThrAlaThrAsnAlaIleAspGlyAlaThrAsnLeuAsn 3424
3223 GACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAA 3260
3425 LysAlaGlnLysAspAlaLeu...LysAlaGlnValThrSerAlaGlnAr 3440
3261 GACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCG 3310
3440 gValAlaAsnValThrSerIleGlnGlnThrAlaAsnGluLeuAsnThrA 3457
3311 GCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3360
3457 laMetGlyGlnLeuGlnHisGlyIleAspAspGluAsnAlaThrLysGln 3473
3361 AGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGC 3410
3474 Thr......GlnLysTyrArgAspAlaGlu...GlnSerLysLysThrAl 3487
3411 CTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCC 3460
3487 aTyrAspGlnAlaValAlaAlaAlaLysAlaIleLeuAsnLysGlnThrG 3504
3461 AACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCC 3510
3504 ly.....Ser 3505
3511 AATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGT 3560
3506 AsnSerAspLysAlaAlaValAspArgAlaLeuGlnGlnValThrSerTh 3522
3561 ACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTT 3610
3522 rLysAspAlaLeuAsn.....GlyAspAlaLysLeuAlaGlu. 3534
3611 GGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGCAAGATTTCCGC 3660
3535 AlaLysAlaAlaAla.....LysGlnAsnLe 3543
3661 GCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCT 3710
3543 uGlyThrLeuAsnHisIleThrAsnAla.....GlnArgThrA 3556
3711 CGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACA 3760
3556 laLeuGluGlyGlnIleAsnGlnAlaThrThrValAspGlyValAsnThr 3572
3761 NCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTT 3810
3573 ValLysThrAsnAlaAsnThrLeuAspGlyAlaMetAsnSerLeuGlnGl 3589
3811 TTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGC 3860
3589 ySerIleAsnAspLysAspAla..... 3596
3861 GGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAA...TCC 3907
3597 ..ThrLeuArgAsnGlnAsnTyrLeuAspAlaAspGluSerLysArgAsn 3612
3908 GCCGCCGCGTGCTGCATTACGGCATTCAGG..... 3937
3613 AlaTyrThrGlnAlaValThrAlaAlaGluGlyIleLeuAsnLysGlnTh 3629
3938 CACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGT...A 3980
3629 rGlyGlyAsnThrSerLysAlaAspValAspAsnAlaLeuAsnThrValT 3646
3981 CATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAA 4030
3646 hrArgAlaLysAlaAlaLeuAsnGlyAlaGluAsnLeuArgAsnThrLys 3662

```
    4031 ACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGC 4080
         ||||||                          :::|||   |||   :::::
    3663 ThrSer.....................AlaThrAsnThrIleAsnGl 3671

    4081 ATTAAGGCAGATTATTCATTCAAACCGGCGCAAC 4114
         :|||   ::::::   :::   ::::::   |||
    3671 yLeuProAsnLeuThrGlnLeuGlnLysAspAsn 3682
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01842

seq_documentation_block:

```
ID   AAB01842 standard; Protein; 998 AA.
XX
AC   AAB01842;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 15 HMW1A protein, SEQ ID NO:59.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52191.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 26A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   998 AA;
```

alignment_scores:

```
           Quality:  240.50            Length:   1196
             Ratio:  0.445               Gaps:     61
 Percent Similarity:  45.151   Percent Identity: 19.900
```

alignment_block:

US-09-303-518D-651 x AAB01842 ..

Align seg 1/1 to: AAB01842 from: 1 to: 998

```
     892 AAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTCTN 941
         |||:::|||:::   |||               |||:::   |||:::
       1 LysGluTrpLeuLeuAspPro...........AspAsnValThrIleGl 13

     942 TTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCC......... 981
                   |||:::|||:::   ::::::   |||
      13 uAlaProSerTyrSerArgGlyAsnAlaGlyIleAspSerGluPheProG 30

     982 .......................AACAACAACGGTACGGGTACGGTA 1005
                                 ::::::||||||       :::
      30 lyGlySerGlyThrLysGluSerProLysThrAsnGlyGluGlnProThr 46

    1006 ACAGAAACCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGT 1055
         :::   |||||||||   :::||||||      ||||||      |||
      47 ValLeuThrAsnGluThrIleSerAsn...TyrLeuLysSerGlyThr.. 61

    1056 CCGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTTACG 1105
                         :::|||   |||   |||:::   :::
      62 .............TrpValMetAsnIleThrAlaLysLysAsnLeu.... 72

    1106 CGGCAGGGGGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAAC 1155
                   |||||||::              :::|||   |||::::::
      73 ........ThrValAsnSer.........SerIleAsnIleGlyAspSer 83

    1156 CTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAACAACAT 1205
              :::|||   :::::::::||||||           ||||||
      84 SerHisLeuIleLeuHisSerGluGlyLys...........AsnAsn.. 95

    1206 CAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGC 1255
                   |||||||::   ::::::   |||:::|||   :::
      96 .............GlyGlyValLysIleLysGluAspIleThrSerAsnG 108

    1256 CTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGAC 1305
            |||              :::|||:::   |||   :::   ::::::
     108 lyGlyAsnLeuThrIleGlnSerGlyGlyTrpValAspValHisLysAsn 124

    1306 AGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAA 1355
           |||
     125 IleThr..................................................... 126

    1356 AATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCT 1405
          :::|||   |||||||||:::::   ||||||             ||||
     127 .LeuGlyThrGlyThrLeuAsnIleThrAlaLys...........GlyS 139

    1406 CGATCAGCGTG...GGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1452
         |||||:::      |||:::||||||
     139 erIleAlaPheGluGlyAsnGlyThr..................... 147

    1453 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAG 1502
         :::|||:::::::::   :::   ::::::|||          ::::::::
     148 GluLysAlaArgAsnAlaSerSerAlaGlnIle.........ThrAlaGl 161

    1503 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1552
         :||||||:::      :::::::|||::::::
     161 nGlyThrIleThrAsnThrGlyAspGlnLys.................. 171
```



```
 618 ysAlaGluIleThrAlaLysAspGly..................... 626

3145 AAAGAACAAGAGCTTTCCGAC...AAACTCGGCAAGGCAGAAGCCAAAAA 3191
                  ||||||      :::|||:::::::::::::
 627 ..............SerAspLeuThrIleGlyAsnThrAsnSer..... 636

3192 ACAGGCGGAAAAAGACAACGCGCAAAGCCTT.................G 3223
       |||:::::::   ||||||:::::::::                   |
 637 ....AlaAspSerThrAsnAlaLysLysValThrPheAsnGlnValLysA 652

3224 ACGCGCTGATTGCGGCCGGG..............CGCGATGCCGCCGAA 3258
     ||:::   |||:::||||||               ::::::
 652 spSerLysIleSerAlaGlyAspHisAsnValThrLeuAsnSerLysVal 668

3259 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGT 3308
     :::|||:::::::   ::::::   :::      ::::::|||:::||||::
 669 GluThrSerGlyAsnThrAspAsnThrGlyAspGlySerGlyAsnAsnAl 685

3309 CGGCATTATGCAGGCGGAG.........GAAGAGAAAAAACGGGTGCAGG 3349
     :|||:::      |||            |||   |||:::::::::   :
 685 aGlyLeuThrIleAlaAlaLysAsnValGluValLysAsnAsnIleThrS 702

3350 CGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCG 3399
     :::::|||          ::::::   :::::::::       |||:::
 702 erAsnLysThrValAsnIleThrAlaSerGluLysLeuThrThrLysAla 718

3400 GNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCA 3449
        :::|||
 719 AspAlaThr.......................................... 721

3450 ACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNA 3499

 721 .............................................. 721

3500 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3549
                               ::::::||||||   ::::::
 722 ............................IleAsnAlaThrThrGlyAsn 728

3550 GTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCG 3599
     |||                      :::|||   ||| ||||||:::
 729 Val.....................GluValThrAla.LysThrGlyA 737

3600 CAACGCNGTTTGGACAA......GCNGCATCCGGNACACCAAACACTACC 3643
               |||:::      ::::::|||||:::   |||  :|||
 737 spIleLys...GlyGluValLysSerThrSerGlyAsnValAsnIleThr 752

3644 GTTCGC............AAGATTTCCGCGCCTACCGCCAACAAACCGAC 3681
     ::::::            :::::::|||      :::::::|||     ||
 753 AlaAsnGlyAspThrLeuAsnValSerAsnValSerGlyAsnAlaValTh 769

3682 CTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCAT 3731
     |   :::   :::      :::   |||:::         ||||||:::|
 769 rIleThrAlaAspLysGlyLysLeuThrThrGln......AlaSerSerS 784

3732 CCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGCA 3781
     ||        :::::::|||      |||   ::::::      ::::::
 784 erIleThrSerAsnAsnGlyGlnThrThrLeuThrAlaLysAspGlySer 800

3782 ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGC 3831
            ::::::   :::|||      :::   ||||||:::::::
 801 IleAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyTh 817

3832 AGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCT 3881
     :   :::|||      :::::::::   ::::::   :::::::::
 817 rLeuThrThrValGluGlySerAsnIleAsnAlaAlaSerGlyThrLeuV 834

3882 NTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCA 3931
        :::|||::::::||||||      :::|||:::
 834 alIleAsnAlaLysAspAlaLysLeuAsnGlyAlaAla............ 846

3932 TTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTAC 3981
               |||   :::   |||:::|||   :::||||||:::
 847 ...SerGlyAspHisThrValValAsnAlaThrAsnAlaSerGlySerGl 862

3982 ATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAA 4031
       |||:::      :::   ||||||:::   :::||||||:::      :
 862 ySerValThrAlaValThrSerSerAsnValAsnIleThrGlyAspLeuS 879

4032 CGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGCA 4081
     ::::::::::
 879 erThrValAsnGly................................... 883

4082 TTAAGGCAGATTATTCATTCAAACCGGCGCAACACA 4117
            :::||||||   :::|||   |||||||||
 884 ...LeuAsnIleIleSerLysAsnGlyArgAsnThr 894
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841

seq_documentation_block:

ID AAB01841 standard; Protein; 1004 AA.
XX
AC AAB01841;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain PMH1 mature HMW2A protein, SEQ ID NO:57.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain PMH1.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52190.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 25A-O; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,

```
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1004 AA;


alignment_scores:
           Quality:  240.00              Length:   1056
             Ratio:  0.481                Gaps:     47
 Percent Similarity:  47.254   Percent Identity: 21.023

alignment_block:
 US-09-303-518D-651 x AAB01841   ..

 Align seg 1/1  to: AAB01841  from: 1  to: 1004

    1369 ACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGG 1418
         |||:::   :::   |||   |||:::|||:::|||||||:::
       5 ThrIleAlaAlaGlyAlaProGlyArgAsnAspGlySerVal........ 18

    1419 CGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAAC 1468
                                      ||||||
      19 ..............................AspAsp.............. 20

    1469 AAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTG 1518
                           :::   :::|||||||||
      21 .................PhePheProThrGlyArgGlyAspAspAlaSer 31

    1519 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1568
         ||||||   ::::::   :::|||||||||
      32 AsnAlaLysThrAsn...HisProAspLys..................... 40

    1569 CGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAA 1618
                                                 :::   |
      41 .....................................ProThrLeuThrA 45

    1619 ATACCGATGAAGGGGCGATGATTGNCNATCATAATGCCACAACAACATCC 1668
         |||||                  :::   :::||||||      ::::::
      45 snThr.................ThrValGluAsnAlaLeuLysAsnAsn 55

    1669 ACC...GTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGTAA 1715
         |||   |||:::||||||:::::::::::::||||||      |||
      56 ThrPheValAsnIleThrAlaLysAsnLysIleThrValAsnSerAspIl 72

    1716 GAATATC..........................AATAGACTTAATTACA 1738
          ||||||                         ::::::   |||   :
      72 eAsnIleLysGlyGlyAlaHisLeuThrLeuTyrSerLysAsnAsnLysL 89

    1739 GCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAA 1788
         ::::::::::      ||||||            :::   |||:::
      89 ysSerSerValLysIleAsnGly............AsnIleThrSerThr 101

    1789 ACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCAC 1838
         |||||||||:::|||                              ||
     102 ThrAsnGlyAsnLeu................................Th 107

    1839 CCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1888
         |      ||||||   ::::::::::::   |||||||||   ::::
     107 rIleTyrSerSerGlyTrpValAspIleHisLysAsnIleThrLeuAsnT 124

    1889 ACGGCAAACTGTTTTTCAGCGGCAGACCGACA.........CCGCACGCC 1929
         ::|||   |||   :::::::::      :::               |||
     124 hrGlyTyrLeuAsnIleThrAlaGlyGlySerValAlaPheGluLysAla 140

    1930 TACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGG 1979
            |||:::   |||         |||:::                |||||
     141 GlyAsnGluLysGlyArgGlnValSerGluSerValIleLysAlaGlnGl 157

    1980 AGAAATCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAA 2029
         |   |||          ::::::   :::   :::   |||:::   :::|
     157 yValIle.........ThrSerGlyValGlyGluGlyPheArgPheAsnA 171

    2030 ATTTCCATATTCAGGGC.......................GGGCAGGCG 2055
         ||      ::::::|||                              ||||||
     171 snValSerLeuAsnGlyValGlyAlaGlyLeuArgPheValGlyGlnLys 187

    2056 GTGATTTCCCGCAAT.....................GTTGCCAAAGTGGA 2084
            ||||||   |||                              :::   ::
     188 AsnIleSerSerAsnSerTrpArgGluAsnThrIleLysAsnArgPheAs 204

    2085 AGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2134
         :|||:::   ::::::|||:::   :::::::::   :::   |||:::
     204 pGlyAsnLeuAsnIleSerGlyLysValAsnValSerMetAspValSerG 221

    2135 CGCATCAAAGCCATACA...ATCTGTACACGTTCGGACTGGACNGGTCTG 2181
               :::   ||||||   |||      |||:::   |||:::
     221 lyThrLysTrpHisThrArgIleAsnGlyArgThrTyrTrpAsnValThr 237

    2182 ACAAATTGTGTCGAANAAANCATTACC..............GACGATAA 2216
         |||      |||   :::      :::                |||   ::
     238 ThrLeuAsnValAlaSerGlySerSerPheAsnLeuSerIleAspAlaSe 254

    2217 AGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNN 2266
         :   |||:::|||   ::::::   |||   ::::::
     254 rGlyIleSerSerGlyAsnGlnAspAspIleThrAsn.............. 266

    2267 CCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAAT 2316
                                                 |||   |||
     267 ..............................ArgGlyLeuAsnGly... 271

    2317 CTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCA 2366
         ::::::   ||||||::::::   :::::::::::::::::::::|||
     272 IleThrPheAsnGlyGluAsnThrPheAsnIleAlaGlnGlySerThrAl 288

    2367 A..................................................... 2367

     288 aAsnPheHisIleLysThrSerValMetThrProLysProAsnSerAsnT 305

    2368 ...........AACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACA 2406
                   |||||||||:::|||::::::|||::::::   ::::::
     305 yrAlaLeuPheAsnGlyAsnIleSerValLeuGlyGlyGlyThrValAsn 321

    2407 TTTAATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGC........ 2448
         |||:::      :::   :::::::::   ::::::||||||
     322 PheGluLeuAsnAlaSerSerSerThrHisThrThrSerGlyAlaIleIl 338

    2449 ....AATGCTTCATTTAATCTAAGCAAC............AACGCCGCAC 2482
             :::   :::||||||::|||:::            |||
     338 eAsnSerGlnAsnPheAsnValSerGlyGlySerLysLeuAsnLeuLysA 355

    2483 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCAT 2532
           :::||||||   :::   :::         |||   |||::::::
     355 laSerGlySerThrAsnThrAlaPheLeuIleLysAsnAsnLeuThrLeu 371

    2533 TCCGCACTCAACGGCAATGTC..........................TC 2555
         :::|||   :::|||||||:::                          ||
     372 AsnAlaThrGlyGlyAsnIleGluIleLysGlnValGluGlyThrAspSe 388

    2556 CCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAC 2605
         |      :::|||:::|||         :::|||      |||   |||
     388 rArgIleGlnLysGlyValValAlaGluGlnAsnIleIlePheGluGlyG 405

    2606 AA......CTCAGCGGCAGCAAGGANACAGCATTACACTTAAAAGACAGC 2649
                  ||||||:::   :::      ::::::|||   :::
     405 lyAsnIleThrLeuGlySerGlnLysAlaProThrGluIleLys...Gly 420

    2650 GAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAA 2699
         :::   |||:::   :::||||||                        ||
     421 AspValThrValLysGlnGlyThr.....................As 429

    2700 CGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGC 2749
```

```
     |||||||:::          :::       :::|||   :::|||
 429 nAlaThrLeuArgSerAlaAsnPheAspAsnHisLysGlyAlaLeuIleV 446

2750 AAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTA 2799
       :::|||
 446 alAsnGly..................................... 448

2800 TTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGAC 2849
       :::|||||||          ::::::    :::       :::|||:::
 449 ...AsnValThrAlaAsnGlyAsnLeuThrAlaAspGlyAspThrIleLy 464

2850 GGTAAACGGCAAATTGAAC...NGTCAAGGAACATTCCGCTTTATGTCGG 2896
       :::::::|||:::|||:::   :::||||||                ::::
 464 sIleLysGlyAsnLeuAspValAlaGlnGly..............AlaL 476

2897 AACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAA 2946
     :::::   |||      :::::::::|||:::::::::
 476 ysPheAsnGlySerThrLysAsnAsnLeuAsnIleThr.........;... 488

2947 GGNACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCT 2996
     ||||||:::|||            |||:::|||::::::   ::::::::
 489 GlyThrPheThr............AsnAsnGlyThrSerIleIleAspIl 501

2997 CGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAA 3046
     :   |||   :::|||      |||      |||          :::   :
 501 eThrGlnGlyValValAsnLeuGly......AsnValThrAsnAspGlyL 516

3047 ACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGT 3096
     ::|||||:::|||   :::   :::      :::   :::      |||
 516 ysLeuAsnIleThrThrHisAlaLysSerGlyGlnLysSerIleIleArg 532

3097 TACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG......CATAATCC 3140
       ::::::|||:::|||:::|||:::::::::::::      :::|||
 533 GlyAspIleIleAsnLysGlnGlyAsnLeuAsnIleThrAspAsnAsnSe 549

3141 GGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAA 3190
            |||   |||::::::   :::::::::::   |||::::::
 549 rAsnAlaGluIleGluIleGlyGlyAsnIleSerGlnLysGluGlyAsnL 566

3191 AACAGGCGGAAAAAGACAACGCGCAA...AGCCTTGACGCGCTGATTGCG 3237
            :::::::|||:::::::::   :::   :::      |||
 566 euThrIleSerSerAspLysValAsnIleThrLysGlnIleThrIleLys 582

3238 GCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCG 3287
     ||||||   |||:::   :::::::::::::|||:::|||:::   |||::
 583 AlaGlyValAspGlyGluSerSerSerSerSerThrAlaSerAspAlaAs 599

3288 GCNGGCA.....................................G 3295
     :   :::                                     :
 599 nLeuThrIleLysThrLysGluLeuThrPheThrAspAsnLeuAsnIleS 616

3296 GCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTG 3345
     ::|||   |||   :::   :::   |||:::::::::::::      :::
 616 erGlyPheAsnLysAlaGluIleThrAlaLysAspAsnSerAspLeuIle 632

3346 ......CAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGA 3389
           :::|||:::::::|||::::::   |||||||||
 633 IleGlyLysAlaSerSerAspAsnSerAsnAlaLysGln........... 645

3390 AACCCGGCCGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGG 3439
                    ::::::|||   :::::::::   ::::::
 646 ..............ValThrPheAspLysValLysAspSerLys.... 655

3440 ATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGC 3489

 655 ..................................................... 655

3490 GACCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCAC 3539
       :::::::::          |||      |||::::::          ||
 656 ...IleSerAlaGlyAsnHisAsnValThrLeuAsnSerLysValGluTh 671

3540 GCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGAC............. 3576
     |   ||||||      ::::::   :::      |||
```

```
 671 rSerAsnSerAspGlySerThrGlyAsnGlySerAspAspAsnAsnIleG 688

3577 .....CGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGC 3621
               :::   |||:::|||                  |||
 688 lyLeuThrIleSerAlaLysAsp..............ValThrValAsn 699

3622 ATCCGGNACACCAAACACTACCGTTCGCAAGATTTCCGCG.......... 3661
       :::   |||:::|||    :::   :::|||||||||
 700 SerAsnIleThrSerHis.LysThrValAsnIleSerAlaSerGluGlyG 716

3662 ...................CCTACCGCCAACAAACCGACCTGCGCCAAA 3691
                                |||   |||      |||   :::
 716 lyIleThrThrLysAlaGlyThrThrIleAsnAlaThrThrGlySerVal 732

3692 TCGGTATGCAGAAAAACC..............TCGGCAGCGGGCGCGTC 3726
     :::|||      ||||||                :::   :::|||   ::
 733 GluValThrAlaLysThrGlyAspIleSerGlyThrIleSerGlyLysTh 749

3727 GGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCAT 3776
     ::::|||         :::||||||      |||          :::|||:
 749 rValSerValThrAlaSerThrGlyAspLeuThrValArgLysAlaAlaT 766

3777 CG...GCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT... 3820
     ::   ::::::   :::|||      |||          :::||||||
 766 hrIleSerValThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnThr 782

3821 ...ACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTT 3867
        |||:::::|||||||||:::   :::|||:::::::::::|||
 783 LeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnValAs 799

3868 AGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGT 3917
       :::|||      |||   :::|||   |||   :::   :::|||:
 799 pLeuSerAla......GlnAspGlySerIleAlaGlyGlnIleSerAlaA 814

3918 GCTGCATTACG..................................... 3928
     ::   :::|||
 814 laAsnValThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySer 830

3929 ........................GCATTCAGGCACGATACCGCGCC 3951
                             |||::::::         |||
 831 AsnIleLysAlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLy 847

3952 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTT 4001
       :::::::::   :::::::::|||||||||:::::::::   :::   :
 847 sLeuAspGlyThrAlaSerGlyAsnArgThrGluValAsnAlaThrAsnA 864

4002 CGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCG 4051
     ::|||   :::      :::::::::|||||||||   |||
 864 laSerGlySerGlySerValThrAlaLysThrSer...SerAsnValAsn 879

4052 GTCTTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTC 4101
     :::      :::   :::|||   ::::::|||   :::||||||   ::
 880 IleThrGlyAspLeuSerThrIleAsnGlyLeu...AsnIleIleSerGl 895

4102 AAACCGGCGCAACACA 4117
     :|||   |||||||||
 895 uAsnGlyArgAsnThr 900
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840

seq_documentation_block:
ID AAB01840 standard; Protein; 1010 AA.
XX
AC AAB01840;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain PMH1 HMW2A protein, SEQ ID NO:55.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

```
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain PMH1.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52189.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 25A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1010 AA;


alignment_scores:
            Quality:  240.00              Length:   1056
              Ratio:  0.481                Gaps:     47
 Percent Similarity:  47.254   Percent Identity: 21.023

alignment_block:
 US-09-303-518D-651 x AAB01840   ..

 Align seg 1/1  to: AAB01840  from: 1  to: 1010

    1369 ACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGG 1418
         |||:::   :::   |||   |||:::|||:::|||||||:::
      11 ThrIleAlaAlaGlyAlaProGlyArgAsnAspGlySerVal..... 24

    1419 CGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAAC 1468
                                  ||||||
      25 ..............................AspAsp............ 26

    1469 AAGCCTTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTG 1518
                                  :::   :::|||||||||
      27 .................PhePheProThrGlyArgGlyAspAspAlaSer 37

    1519 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1568
         ||||||   ::::::   :::|||||||||
      38 AsnAlaLysThrAsn...HisProAspLys..................... 46

    1569 CGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAA 1618
                                                      :::   |
      47 .....................................ProThrLeuThrA 51

    1619 ATACCGATGAAGGGGCGATGATTGNCNATCATAATGCCACAACAACATCC 1668
         |||||                  :::   :::||||||      .::::::
      51 snThr.................ThrValGluAsnAlaLeuLysAsnAsn 61

    1669 ACC...GTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGTAA 1715
         |||   |||:::|||||||:::::::::::::|||||      |||
      62 ThrPheValAsnIleThrAlaLysAsnLysIleThrValAsnSerAspIl 78

    1716 GAATATC.........................AATAGACTTAATTACA 1738
          ||||||                       ::::::   |||   :
      78 eAsnIleLysGlyGlyAlaHisLeuThrLeuTyrSerLysAsnAsnLysL 95

    1739 GCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAA 1788
         ::::::::::      ||||||          :::   |||:::
      95 ysSerSerValLysIleAsnGly.....AsnIleThrSerThr 107

    1789 ACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCAC 1838
         |||||||||:::|||                                ||
     108 ThrAsnGlyAsnLeu.............................Th 113

    1839 CCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1888
         |         ||||||   ::::::::::   |||||||||   ::::
     113 rIleTyrSerSerGlyTrpValAspIleHisLysAsnIleThrLeuAsnT 130

    1889 ACGGCAAACTGTTTTTCAGCGGCAGACCGACA.........CCGCACGCC 1929
         ::|||   |||   :::::::::   :::            |||
     130 hrGlyTyrLeuAsnIleThrAlaGlyGlySerValAlaPheGluLysAla 146

    1930 TACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGG 1979
            |||:::   |||         |||:::              |||||
     147 GlyAsnGluLysGlyArgGlnValSerGluSerValIleLysAlaGlnGl 163

    1980 AGAAATCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAA 2029
         |   |||         ::::::   :::   :::   |||:::   :::|
     163 yValIle.........ThrSerGlyValGlyGluGlyPheArgPheAsnA 177

    2030 ATTTCCATATTCAGGGC.......................GGGCAGGCG 2055
         ||      ::::::|||                      ||||||
     177 snValSerLeuAsnGlyValGlyAlaGlyLeuArgPheValGlyGlnLys 193

    2056 GTGATTTCCCGCAAT.....................GTTGCCAAAGTGGA 2084
            ||||||   |||                             :::   ::
     194 AsnIleSerSerAsnSerTrpArgGluAsnThrIleLysAsnArgPheAs 210

    2085 AGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2134
         :|||:::   :::::|||:::   :::::::::   :::   |||:::
     210 pGlyAsnLeuAsnIleSerGlyLysValAsnValSerMetAspValSerG 227

    2135 CGCATCAAAGCCATACA...ATCTGTACACGTTCGGACTGGACNGGTCTG 2181
               :::   ||||||   |||      |||:::   |||:::
     227 lyThrLysTrpHisThrArgIleAsnGlyArgThrTyrTrpAsnValThr 243

    2182 ACAAATTGTGTCGAANAAANCATTACC..............GACGATAA 2216
         |||      |||   :::      :::               |||   ::
     244 ThrLeuAsnValAlaSerGlySerSerPheAsnLeuSerIleAspAlaSe 260

    2217 AGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNN 2266
         :   |||:::|||   ::::::   |||   ::::::
     260 rGlyIleSerSerGlyAsnGlnAspAspIleThrAsn............. 272
```



```
 789 LeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnValAs 805

3868 AGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGT 3917
      :::|||      |||   :::|||   |||   :::   :::|||:
 805 pLeuSerAla......GlnAspGlySerIleAlaGlyGlnIleSerAlaA 820

3918 GCTGCATTACG..................................... 3928
     ::   :::|||
 820 laAsnValThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySer 836

3929 .........................GCATTCAGGCACGATACCGCGCC 3951
                              |||::::::          |||
 837 AsnIleLysAlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLy 853

3952 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTT 4001
      :::::::::   :::::::::|||||||||:::::::::   :::  :
 853 sLeuAspGlyThrAlaSerGlyAsnArgThrGluValAsnAlaThrAsnA 870

4002 CGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCG 4051
     ::|||   :::     :::::::::|||||||||   |||
 870 laSerGlySerGlySerValThrAlaLysThrSer...SerAsnValAsn 885

4052 GTCTTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTC 4101
     :::      :::   :::|||   ::::::|||   :::||||||   ::
 886 IleThrGlyAspLeuSerThrIleAsnGlyLeu...AsnIleIleSerGl 901

4102 AAACCGGCGCAACACA 4117
     :|||   |||||||||
 901 uAsnGlyArgAsnThr 906
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98256

seq_documentation_block:

```
ID   AAG98256 standard; Protein; 1325 AA.
XX
AC   AAG98256;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Escherichia coli protein sequence SEQ ID NO:304.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
KW   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW   bacterial growth inhibition.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US34419.
XX
PR   23-DEC-1999;   99US-0173005.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen KL,  Zyskind JW;
XX
DR   WPI; 2001-457376/49.
DR   N-PSDB; AAH81312.
XX
PT   Novel nucleic acids encoding proteins required for Escherichia coli
PT   proliferation, useful for screening for antimicrobial agents -
XX
PS   Claim 19; Page 445-448; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in
CC   gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC   of the microorganism, and the manufactured antibiotic is useful for
CC   reducing the activity or level of a gene product required for
CC   proliferation of a microorganism in a subject, specifically humans. The
CC   nucleic acids that inhibit bacterial growth or proliferation can be used
CC   as antisense therapeutics for killing bacteria. In addition to
CC   therapeutic applications, the nucleic acid sequences complementary to
CC   sequences required for proliferation can be used as diagnostic tools.
CC   For example, nucleic acid probes complementary to proliferation-required
CC   sequences that are specific for particular species of microorganisms can
CC   be used as probes to identify particular microorganism species in
CC   clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC   proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC   represent oligonucleotides, which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence   1325 AA;
```

alignment_scores:

```
           Quality:  236.00              Length:    961
             Ratio:  0.524                 Gaps:     50
 Percent Similarity:  46.826   Percent Identity: 20.604
```

alignment_block:

US-09-303-518D-651 x AAG98256 ..

Align seg 1/1 to: AAG98256 from: 1 to: 1325

```
 634 GATTTATCCTACTCCGGCGCATGGTTAATTGGC.............GGCAA 671
     |||::::::|||:::   |||:::|||:::|||             |||
  77 AspValAlaTyrAsp...AlaTyrLeuValGlyTrpTyrGlyThrGlyVa 92

 672 TACACATATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCG 721
         ::::::   :::   ||||||   ::::::::::::::::::   :
  92 lLeuAsnIleLeuAlaGlyGlyAsnAlaSerLeuThrThrIleThrThrS 109

 722 ATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCA 771
     ::|||      ||||||:::                    |||:::::::
 109 erValIleGlyAlaAsnGlu.............AspSerGluGlyThrVal 121

 772 GGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATG 821
     :::      |||:::         :::|||||||:::   ||||||
 122 AsnValLeuGlyGlyThrTrpArgLeuTyrAspSerGlyAsnAsnAlaAr 138

 822 GCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAA 871
         ||||||   |||   |||:::|||
 138 gProLeuAsn...ValGlyGlnSerGly..................... 146

 872 ACGGTTTCCAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGA 921

 146 ................................................. 146

 922 GGCGATACACATACCGTCTNTTTTGAACCGCGCAGTAAC.........GG 962
           |||   |||:::   ::::::   :::::::::         ||
 147 ......ThrGlyThrLeuAsnIleLysGlnLysGlyHisValAspGlyGl 161

 963 ACATTTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAA 1012
     |::::::   :::   |||:::::::::|||:::|||||||||:::
 161 yTyrLeuArgLeuGlySerSerThrGlyGlyValGlyThrValAsnValG 178

1013 CCAACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTG 1062
       :::|||                    :::|||   |||::::::|||
 178 luGlyGluAsp.................SerValLeuThrThrGluLeu 188

1063 TTTGAC.................GAATCTTTGAATGAAACTGATAAAGA 1094
     |||:::                 |||||||||   |||||||||
 189 PheGluIleGlySerTyrGlyThrGlySerLeuAsnIleThrAspLysGl 205

1095 ACCAGTT.........................TACGCGGCAGGGGGTG 1117
         |||                       |||   ||||||
 205 yTyrValThrSerSerIleValAlaIleLeuGlyTyrGlnAlaGly.... 220

1118 TTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATC 1167
```



```
 2680 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCG 2729
      |||:::      ::::::::::::   :::|||:::   |||
  717 GlyLysThrLeuValAsnAspGlyLeuLeuThrIleAlaSer..,..... 730

 2730 CCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCC 2779
       |||   |||   |||:::   ||||||      |||
  731 .HisThrAlaAspGlyVal...ThrGlyMetGlySer.....!.... 741

 2780 GCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAA 2829
                  |||   ::::::::::::   |||
  742 .............SerGluValThrIleAlaAsnPro.....!..... 749

 2830 TCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAAC 2879
             :::||||||   :::   ::::::   |||:::   |||
  750 .........GlyThrLeuAspIleLeuAlaSerThrAsnSerAlaGlyAs 763

 2880 ATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGC 2929
       :::   :::   :::   |||   |||            :::::::::|
  763 pTyrThrLeuThrAsnAlaLeuLysGlyAspGlyLeuMetArgValGlnL 780

 2930 TGGCGGAAAGTTCCGAA...GGNACTTACACCTTGGCGGTCAACAATACC 2976
      ||:::::::|||::::::        :::|||   |||         |||
  780 euSerSerSerAspLysMetPheGlyPheThrHisAla.........Thr 793

 2977 GGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGA 3026
      |||:::|||   :::::::::   ||||||               |||||
  794 GlyThrGluPheAlaGlyValAlaGlnLeu..............LysAs 805

 3027 CAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACG 3076
      |:::                    :::|||||||||:::::::::
  805 pSer.....................ThrPheThrLeuGluArgAsp.... 813

 3077 TCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTC 3126
        :::::::::|||      :::   ::::::::::::|||:::|||
  814 ..AsnThrAlaAlaLeuThrHisAlaMetLeuGlnSerAspSerGluAsn 829

 3127 CGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3159
                  |||   ||||||::::::
  830 ThrThrSerValLysValGlyGluGlnSerIle 840
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01843

seq_documentation_block:

```
ID   AAB01843 standard; Protein; 992 AA.
XX
AC   AAB01843;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 15 mature HMW1A protein, SEQ ID NO:61.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52192.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 26A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   992 AA;
```

alignment_scores:

```
           Quality:  232.50            Length:   1123
             Ratio:  0.454               Gaps:     56
 Percent Similarity:  45.592   Percent Identity: 19.768
```

alignment_block:

US-09-303-518D-651 x AAB01843 ..

Align seg 1/1 to: AAB01843 from: 1 to: 992

```
  691 GGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGA 740
      ||||||   |||:::   |||
   14 GlyAsnAlaGlyIleAspSer.............................. 20

  741 CTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGCGGTTCGCCAA 790
                     :::   |||:::::::|||   :::   ||||||
   21 .............GluPheProGlyGlySerGlyThrLysGluSerPro. 32

  791 TGTTTATTTATGACAAAACAAACAATAAA..................... 819
                  |||||||||::::::
   33 .............LysThrAsnGlyGluGlnProThrValLeuThrAsn 44

  820 ................................TGGCTGCTCAACGGAGT 836
                                      |||::::::|||   ::
   45 GluThrIleSerAsnTyrLeuLysSerGlyThrTrpValMetAsnIleTh 61

  837 TTTACAAACCGGCTACCCTTATTCC........................ 861
      :   :::   :::         :::
   61 rAlaLysLysAsnLeuThrValAsnSerSerIleAsnIleGlyAspSerS 78

  862 ......................GGCAGGGAAAACGGTTTCCAGCTGATA 888
                             |||::::::||||||         |||
   78 erHisLeuIleLeuHisSerGluGlyLysAsnAsnGlyGlyValLysIle 94
```



```
 615 IleThrAlaLys.............................AspGl 620

2295 GCNTGCNNCACTNAANGGCAATCTT...AGTGCAAATGGCGATACACGTT 2341
     |:::           ||||||      ||||||::::::   ::::
 620 ySerAspLeuThrIleGlyAsnThrAsnSerAlaAspSerThrAsnAlaL 637

2342 ATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGC 2391
       |||:::   |||   ::::::::::::::::::|||   :::|||
 637 ysLysValThrPheAsnGlnValLysAspSerLysIleSer...AlaGly 652

2392 AATGCCCAAGCAACATTTAAT...CAAGCCACATTAAACGGCAACNCATC 2438
     :::   ::::::|||:::|||   ::::::      :::||||||:::::
 653 AspHisAsnValThrLeuAsnSerLysValGluThrSerGlyAsnThrAs 669

2439 GGNTTCGGGCAATGCTTCATTTAAT..................... 2463
     :   :::|||::::::|||   |||
 669 pAsnThrGlyAspGlySerGlyAsnAsnAlaGlyLeuThrIleAlaAlaL 686

2464 ..........CTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTT 2502
             :::::::||||||   :::::::|||   ::::::::::::
 686 ysAsnValGluValLysAsnAsnIleThrSerAsnLysThrValAsnIle 702

2503 TCCGACAAC...........GCTAAGGCAAACGTAAGCCATTCCGCACT 2540
     :::   :::           :::||||||:::::::::   :::|||
 703 ThrAlaSerGluLysLeuThrThrLysAlaAspAlaThrIleAsnAlaTh 719

2541 CAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACA 2590
      :::|||||||||:::::::::   |||
 719 rThrGlyAsnValGluValThrAlaLys..................... 728

2591 GCCGCTTTACCGGACAACTCAGCGGCAGC...AAGGANACAGCATTACAC 2637
           ||||||::::::::::|||:::   |||:::|||:::
 729 ........ThrGlyAspIleLysGlyGluValLysSerThrSer...... 740

2638 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2687
                                                ||||||::
 741 .....................................GlyAsnVa 743

2688 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2737
     :|||:::              |||||||||   :::         :::::
 743 lAsnIleThrAlaAsnGlyAspThrLeuAsnValSer.........AsnV 757

2738 CTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGC......CGC 2781
     ::::::|||      :::      :::::::|||      :::
 757 alSerGlyAsnAlaValThrIleThrAlaAspLysGlyLysLeuThrThr 773

2782 CGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATC 2831
     ::::::      |||:::   |||   :::      |||::::::  ::
 774 GlnAlaSerSerSerIleThrSerAsnAsnGlyGlnThrThrLeuThrAl 790

2832 CCGTTTCAACACGCTG..............ACGGTAAACGGCAAATTGA 2866
     ::::   :::::::::            ::::::|||      ||||
 790 aLysAspGlySerIleAlaGlySerIleAsnAlaAlaAsnValThrLeuA 807

2867 ACNGTCAAGGAACATTCCGCTTTATG.....................:.... 2892
     ||:::   ||||||:::      :::
 807 snThrThrGlyThrLeuThrThrValGluGlySerAsnIleAsnAlaAla 823

2893 ......TCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGA 2936
           :::   :::::::::   ::::::   ||||||:::   |||
 824 SerGlyThrLeuValIleAsnAlaLysAspAlaLysLeuAsnGlyAlaAl 840

2937 AAGTTCCGAAGGNACTTACACCTTGGCGGTCAACAATACCGGCAACGAAC 2986
      |||::::::   |||   :::   |||:::|||   :::||||:::
 840 aSerGlyAspHisThrValValAsnAlaThrAsnAlaSerGlySer.... 855

2987 CCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCG 3036
                                       |||:::   :::
 856 ..............................GlySerValThrAlaVal 861

3037 CTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGG 3086
        |||:::|||:::|||:::|||   :::   :::   |||::::::
 862 ThrSerSerAsnValAsnIleThrGlyAspLeuSerThrValAsnGly.. 877

3087 CGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATA 3136
                    ::::::|||   |||:::|||            :::|
 878 ..........LeuAsnIleIleSerLysAsnGly............ArgA 887

3137 ATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCC 3186
     ||   |||                              |||       :::|||
 887 snThrVal.....................ValLeuLysGlyThrGluIle 896

3187 AAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGC 3236
     :::   :::      :::   ::::::   |||:::::::   :::|||
 897 GluValLysTyrIleGlnProGlyValAlaSerValGluGluValIleGl 913

3237 GGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCC 3286
      |||   |||   :::   ||||||:::::::::::::::::      |
 913 uAlaLysArg...ValLeuGluLysValLysAspLeuSerAspGluGluA 929

3287 GGCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAA 3336
     ||   :::   ::::::   :::|||:::      :::
 929 rgGluThrLeuAlaLys...LeuGlyValSerAlaVal..... 940

3337 AAACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGC 3386
     :::   ::::::   ::::::      ::::::      |||:::|||
 941 ArgPheIleGluProAsnAsnThrIleThrValAsnThrGlnAsnGluPh 957

3387 GGAAACCCGGCCGGNTACC 3405
          |||||||||::::::
 957 eThrThrArgProSerSer 963
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:

ID AAB59826 standard; Protein; 1615 AA.
XX
AC AAB59826;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #3 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic

```
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1615 AA;


alignment_scores:
           Quality:  232.00            Length:    733
             Ratio:  0.701               Gaps:     43
 Percent Similarity:  45.157   Percent Identity: 23.738

alignment_block:
 US-09-303-518D-651 x AAB59826   ..

 Align seg 1/1  to: AAB59826  from: 1  to: 1615

    2582 TTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCA 2631
          ||||||   |||:::   |||   :::||||||:::|||      :::||
     552 LeuLysArgAlaSerArgProAlaSerSerAlaCysAlaIleProLysHi 568

    2632 T.....TACACTT 2639
         |     ::::::
     568 sThrArgProAlaLysAlaAsnProLysProSerValAlaArgTrpAlaT 585

    2640 AAAAGACAGCGAATGGACGC......TGCCGTCAGGCACGGAATTAGGCA 2683
           :::|||::::::::::|||      |||   :::      |||   :::
     585 rpArgThrSerThrSerArgProTrpCysSerArgArgThrAsnSerSer 601

    2684 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2733
               :::   |||||||||         :::|||||||      ::: |
     602 ......SerAlaThrThrPro.........LysIleProThrCysSer.A 613

    2734 GATGC.....TGCAGGCGC 2747
         |||||          ||| |||||
     613 rgCysIleProAsnCysProThrTrpProCysArgThrThrCys.GlyAl 629

    2748 GCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCC 2797
         |   |||               |||   ||||||||||||   :::::
     629 aThrThrArgArgSerArgProThrArgArgArgArgSerMetAsnThrG 646

    2798 TATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTG 2847
               ::::::     :::||||||| :::|||:::|||
     646 lySerArgIleAlaCysArgAlaSerVal.SerProIleSer........ 659

    2848 ACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGA 2897
                              :::::::::   ||||||   |||   :
     660 .....IleArgGlnThrSerAlaAlaCysIleA 669

    2898 ACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAG 2947
         ::|||:::|||                  ||||||:::
     669 rgSerAlaAla.....TrpArgArg..... 675

    2948 GNACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTC 2997
                    |||   ::::::|||:::
     676 ..............ArgProSerLeuProVal..... 681

    2998 GATCAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAA 3047
                                  ||||||     ||| |||
     682 ..........................ThrThrAlaSerCys.Arg... 687

    3048 CCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTT 3097
                  |||:::|||   :::   ::: :::|||:::||||||
     688 ............ProThrLysProSerTrpLys.ThrGlyCysTrpArg. 699

    3098 ACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAA 3147
         ::::::||||||   |||:::   |||
     700 AlaSerSerSerProLysSerIleSerProLys..... 710

    3148 GAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGC 3197
                   |||      ::::::|||   :::|||   :::   |
     711 ............ProArgProThrCysArgProSerProGlyThrAlaA 723

    3198 GGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCG 3247
         ||:::   :::|||   :::                      |||   :::
     723 rgArgValSerThrThrSer.....ProArgSer 732

    3248 ATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGC 3297
                  :::|||   ::::::      :::|||   :::   :::||
     733 ThrThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArgAl 749

    3298 GGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCA 3347
         ||||:::      |||   |||   |||   :::||||||::::::   |
     749 aGlyArg......AlaGlyCysAlaArgSerSerArgLysThrSer...A 763

    3348 GGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGC 3397
         ||   |||::::::|||      |||   |||:::::::::::::|||
     763 rgProIleArgSerAla......ArgProSerCysSerLysSerProThr 777

    3398 CGGNTACCA......CCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGAT 3441
                        ||||||||||||||||   |||:::   :::
     778 SerValSerAlaPheProProSerProAlaArgAlaSerArgThrArgCy 794

    3442 TTGCCGCAACCGCAGCCCCAAC......CGCAACCTCAACCCCAACCGCA 3485
               ||||||      |||:::         :::   :::   :::||||:
     794 sArgArgAsnSerLeuProSerSerValThrArgSerSerAlaThrArgA 811

    3486 GCGCGACCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCG 3535
         ::||||||                |||   :::      ::::::   |||
     811 laAlaThrProArgArgLysThrProCysCysGlyArgThrThrArgPro 827

    3536 CCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGG..... 3574
         |||   ||||||      |||      |||::::::|||
     828 ProSerSerThrArgAsnSerSer...ArgAlaThrTrpMetArgTrpAs 843

    3574 ..... 3574

     843 nSerSerArgTrpAsnValArgPheProSerMetAlaProAlaSerArgA 860

    3575 .....ACCGCGTGT......TTGCCG 3589
                              |||   |||         |||
     860 laProThrAlaLysSerSerArgGlyArgThrIleCysSerSerSerPro 876

    3590 AAGACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGNACACCAAACAC 3639
         ::::::|||   |||      :::::::::   :::   ||||||:::
     877 SerAlaAlaProThrProArgAlaArgThrProAlaThrThrProThrPr 893

    3640 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3689
          :::   |||:::   |||:::   :::      :::|||      ::::
     893 oSerSerArgGlnProSerGlySerAlaArgProSerProProSerSerS 910

    3690 AATCGGTATGCAGAAAAACCTCG.....GCAGCGGGCGCGTCGGCA 3730
         ::::::::   |||:::|||:::         :::||||||   |||:::
     910 erAlaIleProArgArgThrAlaArgArgArgCysAlaGlyPheSerSer 926

    3731 TCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGC 3780
         :::              ::::::   :::   :::|||||||||   |||||
     927 AlaSerAlaThrAspSerAlaIleArgArgSerSerThrThrArgSerAl 943

    3781 A...ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCAT 3827
         |   :::|||::::::   |||:::         |||::::::||||||:
     943 aArgSerArgArgAsnThrProSer.....SerAlaSerThrAlaT 957

    3828 CGGCA.....GGTTCGACATCGGCATCAGCA 3853
         ::|||                |||||||||   :::::::::
     957 hrAlaProProThrArgLysProThrThrGlySerThrCysCysAlaCys 973

    3854 CGG...GCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAG.. 3898
         |||   |||   ::::::||||||         :::   :::   :::
     974 ArgProAlaSerThrValAlaAlaArgArgLysLysProValArgLysVa 990

    3899 ....GCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATA 3944
             |||:::|||:::      |||      :::   |||      |
     990 lAlaAlaGlnSerSerArgProSerCysTrpLysSer...ArgSerMetT 1006

    3945 CCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGC 3994
```

```
      |||||   ::::::         :::   |||      ||||||:::|||
 1006 hrAlaThrThrGlyArgThrProThrCysAsn......SerAlaArgArg 1020

 3995 GCTATTTCGTCCAAAAAG.....................CGGATTACCGC 4023
        :::   |||:::::                       |||:::::||
 1021 ProValIleSerArgArgSerProSerArgMetPheGlyArgLeuSerAl 1037

 4024 TACG..............AAAACGTCAATATCGCCACCCCCGGTCTTGC 4058
      |:::              :::||||||:::|||   |||   :::   |
 1037 aSerSerIleAsnMetArgSerThrSerValSerAlaProArgThrCysA 1054

 4059 GTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAACCGG 4108
      ||:::|||:::::::::   :::   |||   :::      ||| |:
 1054 rgAlaThrSerSerSerAlaSerCysArgCysLeu......SerCysPro 1068

 4109 CGCAACACATNTCCATCACNCCTTATTTNAGCCTGTCCTATACCGATGCC 4158
      ::::::|||:::   :::                  |||   |||   ||
 1069 GlnSerThrThrAlaAlaTrpAsnSerGlyTrpThrProAlaProCysPr 1085

 4159 G.....................CTTCGGGCAAAGTCCGAACACGCGTCAA 4187
      |                      |||::::::   :::
 1085 oSerSerProMetAlaGlyThrThrArgSerArgArgSerSerArgArgT 1102

 4188 TACCGCNGTAT............TGGCTC..................... 4204
        |||::::::         |||
 1102 hrProSerTrpProSerArgAsnTrpTyrSerArgArgArgAsnThrPro 1118

 4205 ...............................AGGATTTCGGCAAAA.. 4219
                                     ::::::|||   |||
 1119 SerSerAsnSerAlaLysArgArgArgThrGlyLysValSerArgLysCy 1135

 4220 ..............................................C 4220
                                                    |
 1135 sAlaSerThrSerSerGlyArgArgSerGlyAlaThrThrMetIleThrP 1152

 4221 CCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGT 4270
      ||:::|||   :::   |||   |||   ||||||:::::::::::::|||
 1152 roThrValSerSerProAlaSerThrArgLysSerSerAlaAlaLysCys 1168

 4271 .........CCNTCCACGCTGCCGCCGCCAAAGGNCCGC........... 4300
              |||:::||||||         :::   |||
 1169 AlaArgSerProThrThrLeuValValArgSerCysArgLeuValArgLe 1185

 4301 ..........AACTGGAAGCGCAACACAGCGCGGGCA 4327
               :::||||||         ||||||||
 1185 uSerAlaCysThrTrpLysSerValArgAlaArgAla 1197
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834

seq_documentation_block:

```
ID   AAB01834 standard; Protein; 1101 AA.
XX
AC   AAB01834;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52183.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 22A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1101 AA;
```

alignment_scores:

```
           Quality:  231.00              Length:   1008
             Ratio:  0.456                 Gaps:     55
 Percent Similarity:  50.298   Percent Identity: 21.329
```

alignment_block:

```
 US-09-303-518D-651 x AAB01834   ..

 Align seg 1/1  to: AAB01834  from: 1  to: 1101

     551 ATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 600
         |||   ||||||:::|||         :::      |||   :::|||
      77 IleAsnIleLysGluAsnSerHisLeuIleLeuTrpSerGluArgAspGl 93

     601 CACTATTGGCGTTATGATGATGACAAACAC..............GGCG 634
          :::   ||||||::::::      |||                 |||:
      93 yAsnSerGlyValGlnIleAspGlyAsnIleThrSerAlaThrGlyGlyS 110

     635 ATTTATCC...TACTCCGGCGCATGGTTAATTGGCGGCAATACACATATG 681
         ::|||:::   ||||||::::::|||:::         ::::::|||
     110 erLeuThrValTyrSerSerGlyTrpVal.........AspValHisLys 123

     682 CAGGGTTGGGGAAATAATGGC.........GTANTTAGTTTGAGCGGCGA 722
         :::         |||:::|||         :::::::::   ||||||||
     124 AsnIleThrLeuAsnSerGlyTyrLeuAsnIleThrThrLysSerGlyAs 140

     723 TGTG......CGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTG 766
         ||||      :::::::::||||||         :::   |||:::|||
```



```
2360 CCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTT 2409
     ::|||   |||:::::::::::::
 660 alThrLeuAsnSerLysValGlu..................................... 667

2410 AATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAAT........ 2451
              |||   |||||||||   ::::::::::::|||
 668 .........ThrSerAsnGlyAsnAsnAspAlaGluSerAsnAsnGlyAs 681

2452 ..........GCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGC. 2490
                 ::::::|||   :::|||   ::::::::::|||:::
 681 pGlyThrSerLeuThrIleAsnAlaLysAsnIleThrValAsnAsnAsnI 698

2491 ...................AGTCTGACGCTTTCCGACAAC......GCT 2514
                         :::::::|||   |||:::|||      :::
 698 leThrSerHisLysThrValAsnIleThrAlaSerGluAsnValThrThr 714

2515 AAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGA 2564
     ||||||:::::::::   :::|||   :::|||:::|||:::::::::
 715 LysAlaGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAl 731

2565 TAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGCG 2614
      |||                                |||||||:::::::::|
 731 aLys..............................ThrGlyAspIleLysG 738

2615 GCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCG 2664
     ||:::           ::::::      :::::::::   ||||||
 738 lyLys............ValGluSerThrSerGlySerValThrLeuThr 750

2665 TCAGGCACGGAA.........TTAGGCAATTTAAACCTTGACAACGCCAC 2705
     :::      |||         :::::::|||::::::   ::::::::::||
 751 AlaThrGlyGluAlaLeuAlaValSerAsnIleSerGlyAsnThrValTh 767

2706 CATTACACTCAAT.....................................T 2719
     |||||||   |||                                     |
 767 rIleThrAlaAsnLysGlyLysLeuThrThrGlnAlaGlySerThrValS 784

2720 CCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGAC 2769
     |||||   :::      :::|||::::::|||:::|||   :::|||
 784 erAlaIleAsnGlyValThrAlaSerSerGlnSerGlyAspIleSerGly 800

2770 ACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAAC 2819
     |||                                  ::::::   : ||
 801 Thr.............................IleSerGlyAsnTh 806

2820 TTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACN 2869
     |         |||   ::::::   ||||||::::::::
 806 rValLysValSerAlaIleGlyAspLeuThrThrLysSerGly..... 820

2870 GTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGAC 2919
                          ||||||:::
 821 ......................SerGluIle................. 823

2920 AAATTGAAGCTGGCGGAA............AGTTCCGAAGGNACTTAC.. 2955
     |||   |||   :::|||           |||:::   ||||||
 824 LysAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGl 840

2956 ....ACCTTG.........GCGGTCAAC.........AATACCGGCAAC. 2982
         |||:::         |||||||||         |||||||||:::
 840 yGlyThrIleSerGlyAsnAlaValAsnValThrAlaAsnThrGlyAspL 857

2983 ...........GAACCCGTAAGCCTCGAT.................CAA 3003
                :::   ::::::::::|||
 857 euThrValGluAspAlaAlaLysIleAspAlaThrGlyGlyAlaAlaThr 873

3004 TTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAA 3053
     ||||||:::::: :::|||||||   |||:::::::::   :::::::::
 874 LeuThrAlaThr.SerGlyLysLeuThrThrLysAlaSerSerSerIleT 890

3054 TTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAAC 3103
       |||                              |||::::::   :::|||
 890 hrSer.........................AlaAsnAsnGlnValAsn 897

3104 TCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3153
```

```
          |||||||||   :::|||   :::      |||::::::   |||
 898 LeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsnVa 914

3154 GAGCTTTCCGACAAACTCGGCA 3175
        ::::::   |||:::::::|||
 914 lThrLeuAsnThrThrGlyAla 921
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW04505

seq_documentation_block:

ID AAW04505 standard; Protein; 1992 AA.
XX
AC AAW04505;
XX
DT 25-JAN-1997 (first entry)
XX
DE Moraxella 200 kDa outer membrane protein.
XX
KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
KW diagnosis.
XX
OS Mycobacterium catarrhalis strain 4223.
XX
PN WO9634960-A1.
XX
PD 07-NOV-1996.
XX
PF 29-APR-1996; 96WO-CA00264.
XX
PR 26-MAR-1996; 96US-0621944.
PR 01-MAY-1995; 95US-0431718.
PR 07-JUN-1995; 95US-0478370.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX
DR WPI; 1996-506162/50.
DR N-PSDB; AAT38740.
XX
PT Moraxella outer membrane protein - useful as immunogen in protective
PT vaccine and for diagnosis
XX
PS Claim 14; Fig 6; 109pp; English.
XX
CC An approx. 200 kDa outer membrane protein (AAW04505) can be
CC isolated from Moraxella catarrhalis otitis media strain 4223
CC by electroelution, or expressed from a gene (see also AAT38740)
CC obtd. from a strain 4223 genomic library. Natural or recombinant
CC outer membrane protein is useful as an immunogen to protect
CC against infection by Moraxella, esp. M. catarrhalis. It can
CC also be used to detect antibodies, esp. for differential diagnosis
CC between bacteria that cause similar symptoms, and also useful as
CC a carrier for other antigens and used to raise antitumour
CC antibodies for conjugation to therapeutic agents.
XX
SQ Sequence 1992 AA;

alignment_scores:

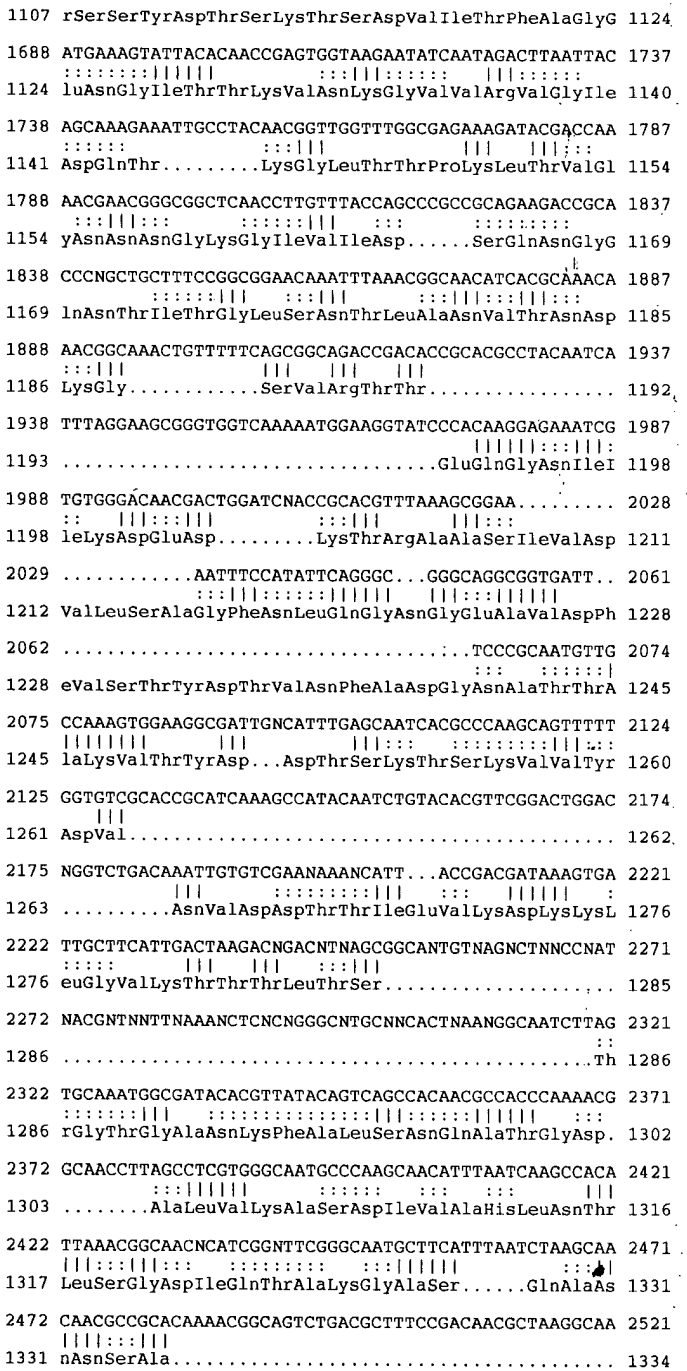
Quality:	230.00	Length:	1653
Ratio:	0.316	Gaps:	72
Percent Similarity:	44.102	Percent Identity:	18.391

alignment_block:
US-09-303-518D-651 x AAW04505 ..

Align seg 1/1 to: AAW04505 from: 1 to: 1992

```
 163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC.....GG 200
     ::|||   :::::::|||   |||      ::::::        ||
 602 LysAspThrThrLysAsnAlaGlyAlaValSerIleLeuLysLeuLysGl 618
```




1107 rSerSerTyrAspThrSerLysThrSerAspValIleThrPheAlaGlyG 1124
1688 ATGAAAGTATTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTAC 1737
1124 luAsnGlyIleThrThrLysValAsnLysGlyValValArgValGlyIle 1140
1738 AGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAA 1787
1141 AspGlnThr.....LysGlyLeuThrThrProLysLeuThrValGl 1154
1788 AACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCA 1837
1154 yAsnAsnAsnGlyLysGlyIleValIleAsp.....SerGlnAsnGlyG 1169
1838 CCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACA 1887
1169 lnAsnThrIleThrGlyLeuSerAsnThrLeuAlaAsnValThrAsnAsp 1185
1888 AACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCA 1937
1186 LysGly.....SerValArgThrThr..... 1192
1938 TTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCG 1987
1193 GluGlnGlyAsnIleI 1198
1988 TGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAA..... 2028
1198 leLysAspGluAsp.....LysThrArgAlaAlaSerIleValAsp 1211
2029 AATTTCCATATTCAGGGC...GGGCAGGCGGTGATT.. 2061
1212 ValLeuSerAlaGlyPheAsnLeuGlnGlyAsnGlyGluAlaValAspPh 1228
2062 TCCCGCAATGTTG 2074
1228 eValSerThrTyrAspThrValAsnPheAlaAspGlyAsnAlaThrThrA 1245
2075 CCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTT 2124
1245 laLysValThrTyrAsp...AspThrSerLysThrSerLysValValTyr 1260
2125 GGTGTCGCACCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGAC 2174
1261 AspVal..... 1262
2175 NGGTCTGACAAATTGTGTCGAANAAANCATT...ACCGACGATAAAGTGA 2221
1263 AsnValAspAspThrThrIleGluValLysAspLysLysL 1276
2222 TTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNAT 2271
1276 euGlyValLysThrThrThrLeuThrSer..... 1285
2272 NACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAG 2321
1286 Th 1286
2322 TGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACG 2371
1286 rGlyThrGlyAlaAsnLysPheAlaLeuSerAsnGlnAlaThrGlyAsp. 1302
2372 GCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACA 2421
1303 AlaLeuValLysAlaSerAspIleValAlaHisLeuAsnThr 1316
2422 TTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAA 2471
1317 LeuSerGlyAspIleGlnThrAlaLysGlyAlaSer.....GlnAlaAs 1331
2472 CAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAA 2521
1331 nAsnSerAla..... 1334
2522 ACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCA 2571
1335 GlyTyrValAspAlaAspGlyAsnLysValIleTyrAspSerThr 1349
2572 GTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAA 2621
1350 AspAsnLysTyrTyrGlnAlaLysAsnAspGlyThrValAspLysThrLy 1366
2622 GGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCA 2671
1366 sGluValAlaLysAspLysLeuValAlaGlnAlaGlnThrProAspGlyT 1383
2672 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2721
1383 hr...LeuAlaGlnMetAsnValLysSer.....ValIleAsnLys 1395
2722 GCCTATCGCCACGATGCTGCAGGCGCGCAA..... 2751
1396 GluGlnValAsnAspAlaAsnLysLysGlnGlyIleAsnGluAspAsnAl 1412
2752 ACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTT 2785
1412 aPheValLysGlyLeuGluLysAlaAlaSerAspAsnLysThrLysAsnA 1429
2786 CGCGCCGTTCCCTA.....TTATCCGTTACACCGCCAACTTCGGTAGAA 2829
1429 laAlaValThrValGlyAspLeuAsnAlaValAlaGlnThrProLeuThr 1445
2830 TCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAAC 2879
1446 PheAlaGlyAspThrGlyThrThrAlaLysLysLeuGlyGluThrLeuTh 1462
2880 ATTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGC 2929
1462 rIleLys.....GlyGlyGlnThrAspThrAsnLysL 1473
2930 TGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCAACAATACCGGC 2979
1473 euThrAspAsnAsnIleGlyValVal..... 1481
2980 AACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAA 3029
1482 AlaGlyThrAspGlyPheThrValLysLeuAlaLysAsp.. 1494
3030 CAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCG 3079
1495 LeuThrAsnLeuAsn.....SerValA 1502
3080 ATGCCGGCGCGTGGCGT.....TACCAACTCATCCGCAAA 3114
1502 snAlaGlyGlyThrLysIleAspAspLysGlyValSerPheValAspSer 1518
3115 GACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGA 3164
1519 SerGlyGlnAlaLysAlaAsnThrProValLeuSerAlaAsnGlyLeuAs 1535
3165 CAAACTCGGCAAG.....GCAGAAGCCAAAAAACAGGCGG 3199
1535 pLeuGlyGlyLysValIleSerAsnValGlyLysGlyThrLysAspThrA 1552
3200 AAAAAGACAACGCGCAAAGCCTTGAC.....GCGCTG 3231
1552 spAlaAlaAsnValGlnGlnLeuAsnGluValArgAsnLeuLeuGlyLeu 1568
3232 ATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAA.....AG 3269
1569 GlyAsnAlaGlyAsnAspAsnAlaAspGlyAsnGlnValAsnIleAlaAs 1585
3270 CGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATGC 3319
1585 pIleLysLysAspProAsnSerGlySerSerSerAsnArgThrValIleL 1602

```
3320 AGGCG............................GAGGAAGAGAAAAAA 3339
     ::|||                           :::   ||||||
1602 ysAlaGlyThrValLeuGlyGlyLysGlyAsnAsnAspThrGluLysLeu 1618

3340 ..............CGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAA 3374
                   :::|||   :::|||||||||||:::
1619 AlaThrGlyGlyIleGlnValGlyValAspLysAspGly........... 1631

3375 ACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGCCTTCCCCCGCGCCC 3424

1631 ................................................. 1631

3425 GCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAA 3474
       |||:::   ||||||   :::         :::   |||   :::
1632 ..AsnAlaAsnGlyAspLeuSerAsnValTrpValLysThrGln...Lys 1646

3475 CCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGCGGTTTGAG 3524
             ::::::   |||   :::   |||      :::|||   ::
1647 AspGlySerLysLysAlaLeuLeuAlaThrTyrAsnAlaAlaGlyGlnTh 1663

3525 TGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGG 3574
     :::::::      |||   |||   :::   |||:::             |
1663 rAsnTyr...LeuThrAsnAsnProAlaGluAlaIle............A 1675

3575 ACCGCGTGTTTGCCGAA............................... 3591
     |||||:::      :::
1675 spArgIleAsnGluGlnGlyIleArgPhePheHisValAsnAspGlyAsn 1691

3592 ..............GACCGCCGCAACGCNGTTTGGACAAGCNGCATCCG 3626
                   :::   ||||||:::::   :::|||:::
1692 GlnGluProValValGlnGlyArgAsnGlyIleAspSerSerAla..... 1706

3627 GNACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAA 3676
      :::   ||||||      :::      |||:::|||      :::
1707 .SerGlyLysHisSerValAlaIleGlyPheGlnAlaLysAlaAspGlyG 1723

3677 CCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTC 3726
                  ||||||   |||   :::   |||:::      :::
1723 luAlaAlaValAlaIleGlyArgGlnThrGlnAlaGlyAsnGlnSerIle 1739

3727 GGCATCCTGTTTTCGCACAACCGGACC...GAAAACANCTTCGACGACGG 3773
     :::|||      :::   :::   |||   ::::::::::::      ||
1740 AlaIleGlyAspAsnAlaGlnAlaThrGlyAspGlnSerIleAlaIleGl 1756

3774 CATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACG 3823
     |   ||||||   :::         |||:::::::::::|||:::'   :
1756 yThrGlyAsnValValAlaGlyLysHisSerGlyAlaIleGlyAspProS 1773

3824 GCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT......TTT 3867
     ::      :::   |||   :::   |||:::|||   :::      |||
1773 erThrValLysAlaAspAsnSerTyrSerValGlyAsnAsnAsnGlnPhe 1789

3868 AGCAGCGGCANTCTNTCAGAC......GGCATCGGAGGCAAAATCCGC.. 3909
     ::::::::::::   :::|||      |||:::|||:::::::|||
1790 ThrAspAlaThrGlnThrAspValPheGlyValGlyAsnAsnIleThrVa 1806

3910 ..........................................CGCCGCG 3916

1806 lThrGluSerAsnSerValAlaLeuGlySerAsnSerAlaIleSerAlaG 1823

3917 TGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGC...... 3960
          |||   |||   ||||||:::         |||   :::
1823 lyThrHisAlaGlyThrGlnAlaLysLysSerAspGlyThrAlaGlyThr 1839

3961 ................................GGATTCGGCATCGAACC 3977
                                     ||||||:::   :::
1840 ThrThrThrAlaGlyAlaThrGlyThrValLysGlyPheAlaGlyGlnTh 1856

3978 GTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACG 4027
         :::||||||:::         ::::::   |||:::   |||   :
1856 rAlaValGlyAlaValSerValGlyAlaSerGlyAlaGluArgArgIleG 1873

4028 AAAACGTCAATATC.................................. 4041
```

```
     ::||||||
1873 lnAsnValAlaAlaGlyGluValSerAlaThrSerThrAspAlaValAsn 1889

4042 ..........................................GCCACCCCCGG 4052
                                               ||||||
1890 GlySerGlnLeuTyrLysAlaThrGlnSerIleAlaAsnAlaThrAsnGl 1906

4053 TCTTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCA 4102
      |||      :::   ::::::         ||||||:::   :::::::
1906 uLeuAspHisArgIleHisGlnAsnGluAsnLysAlaAsnAlaGlyIleS 1923

4103 AACCGGCGCAACACATNTCCATCACNCCTTATTTNAGCCTGTCCTATACC 4152
     ::   |||:::      :::      |||   :::   :::
1923 erSerAlaMetAlaMetAlaSerMetProGlnAlaTyrIleProGlyArg 1939

4153 GATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCNGTATTGGC 4202
     :::   ::::::|||   :::   |||:::   :::   :::::::::||
1940 SerMetValThrGlyGlyIleAlaThrHisAsnGlyGlnGlyAlaValAl 1956

4203 TCAGGATTTCGGCAAA...ACCCGCAGTGCGGAATGG............G 4237
     |      ::::::|||   :::   :::::::::|||           :
1956 aValGlyLeuSerLysLeuSerAspAsnGlyGlnTrpValPheLysIleA 1973

4238 GCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCNTCCACGCTGCCGCC 4287
     ::   :::|||:::   :::|||               |||::::::|||
1973 snGlySerAlaAspThrGlnGly..............HisValGlyAla 1984

4288 GCCAAAGGN 4296
     |||   |||
1985 AlaValGly 1987
```